

FORM PTO-1390 (REV. 11-2000)		U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE	ATTORNEY'S DOCKET NUMBER R-341894
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371			U.S. APPLICATION NO. (If known, see 37 CFR 1.5) 09/830902
INTERNATIONAL APPLICATION NO. PCT/FR00/02433	INTERNATIONAL FILING DATE 4 September 2000	PRIORITY DATE CLAIMED 3 September 1999	
TITLE OF INVENTION CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL DOMINANT SPASTIC PARAPLEGIA			
APPLICANT(S) FOR DO/EO/US Jean Weissenbach, Jamile Hazan			
Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:			
<ol style="list-style-type: none"> 1. <input checked="" type="checkbox"/> This is a FIRST submission of items concerning a filing under 35 U.S.C. 371. 2. <input type="checkbox"/> This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371. 3. <input type="checkbox"/> This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below. 4. <input checked="" type="checkbox"/> The US has been elected by the expiration of 19 months from the priority date (Article 31). 5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2)) <ol style="list-style-type: none"> a. <input checked="" type="checkbox"/> is attached hereto (required only if not communicated by the International Bureau). b. <input checked="" type="checkbox"/> has been communicated by the International Bureau. c. <input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US). 6. <input checked="" type="checkbox"/> An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)). <ol style="list-style-type: none"> a. <input checked="" type="checkbox"/> is attached hereto. b. <input type="checkbox"/> has been previously submitted under 35 U.S.C. 154(d)(4). 7. <input type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)) <ol style="list-style-type: none"> a. <input type="checkbox"/> are attached hereto (required only if not communicated by the International Bureau). b. <input type="checkbox"/> have been communicated by the International Bureau. c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired. d. <input type="checkbox"/> have not been made and will not be made. 8. <input type="checkbox"/> An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)). 9. <input type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)). 10. <input type="checkbox"/> An English language translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)). 			
Items 11 to 20 below concern document(s) or information included:			
<ol style="list-style-type: none"> 11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98. 12. <input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included. 13. <input checked="" type="checkbox"/> A FIRST preliminary amendment. 14. <input type="checkbox"/> A SECOND or SUBSEQUENT preliminary amendment. 15. <input checked="" type="checkbox"/> A substitute specification. 16. <input type="checkbox"/> A change of power of attorney and/or address letter. 17. <input checked="" type="checkbox"/> A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825. 18. <input type="checkbox"/> A second copy of the published international application under 35 U.S.C. 154(d)(4). 19. <input type="checkbox"/> A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4). 20. <input checked="" type="checkbox"/> Other items or information: See Attachment A. 			

U.S. APPLICATION NO. (if known, see 37 CFR 1.51) 09/830902		INTERNATIONAL APPLICATION NO. PCT/FR00/02433		ATTORNEY'S DOCKET NUMBER R-341894	
--	--	--	--	---	--

21. <input checked="" type="checkbox"/> The following fees are submitted: BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)): Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO \$1000.00 International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO \$860.00 International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO \$710.00 International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) \$690.00 International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) \$100.00 ENTER APPROPRIATE BASIC FEE AMOUNT =				CALCULATIONS PTO USE ONLY	
				\$ 860.00	
Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input checked="" type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				\$ 130.00	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	\$	
Total claims	40 - 20 =	20	x \$18.00	\$ 360.00	
Independent claims	28 - 3 =	25	x \$80.00	\$ 2,000.00	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$270.00	\$	
TOTAL OF ABOVE CALCULATIONS =				\$ 3,350.00	
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				\$	
SUBTOTAL =				\$ 3,350.00	
Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				\$	
TOTAL NATIONAL FEE =				\$ 3,350.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property +				\$	
TOTAL FEES ENCLOSED =				\$ 3,350.00	
				Amount to be refunded:	\$
				charged:	\$

a. ☒ A check in the amount of \$ 3,350.00 to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. _____ in the amount of \$ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.


c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 13-2100. A duplicate copy of this sheet is enclosed.

d. ☐ Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Joseph Krieger
 Mason, Kolehmainen, Rathburn & Wyss
 853 Sanders Road, #330
 Northbrook, Illinois 60062


 SIGNATURE
Joseph Krieger
 NAME
25,595
 REGISTRATION NUMBER

Attachment A

FORM PTO-1390 (REV. 11-2000)		U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE	ATTORNEY'S DOCKET NUMBER R-341894
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371			U.S. APPLICATION NO. (If known, see 37 CFR 1.5) 09/830902
INTERNATIONAL APPLICATION NO. PCT/FR00/02433	INTERNATIONAL FILING DATE 4 September 2000	PRIORITY DATE CLAIMED 3 September 1999	
TITLE OF INVENTION CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL DOMINANT SPASTIC PARAPLEGIA			
APPLICANT(S) FOR DO/EO/US Jean Weissenbach, Jamile Hazan			

Item 20 - Other Items or Information:

- (a) Certificate Of Mailing By Express Mail For National Phase
Application Of International Application No. PCT/FR00/02433
- (b) Seven (7) Drawing sheets
- (c) Identification of Inventors
- (d) Application Data Sheet
- (e) Copies of
 - (i) PCT Request (Form PCT/RO/101)
 - (ii) PCT Notification of Receipt of Record Copy (Form PCT/IB/301)
 - (iii) PCT Notification Concerning Submission or Transmittal of
Priority Document (Form PCT/IB/304)
 - (iv) PCT Notice Informing Applicant of the Communication of the
International Application to the Designated Offices (Form
PCT/IB/308)

09/830902-101901

09/830902

JC08 Rec'd PCT/PTO 02 MAY 2001

**CERTIFICATE OF MAILING BY EXPRESS MAIL
FOR NATIONAL PHASE APPLICATION OF
INTERNATIONAL APPLICATION NO. PCT/FR00/02433**

"Express Mail" Mailing Label No. **EL713287403US.**

I hereby certify that the Transmittal Letter to the United States Designated/Elected Office (DO/EO/US) Concerning a Filing Under 35 U.S.C. 371 and the documents referred to therein and the fee referred to therein are being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated below and is addressed to the Commissioner for Patents, Washington, D.C. 20231


Signature

05/02/2001

Date of Deposit

Joseph Krieger

Typed/printed name of person signing

09/830902

09/830902

JC08 Rec'd PCT/PTO 02 MAY 2001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application No. :

U.S. National Serial No. :

Filed :

PCT International Application No. : PCT/FR00/02433

VERIFICATION OF A TRANSLATION

I, the below named translator, hereby declare that:

My name and post office address are as stated below;

That I am knowledgeable in the French language in which the below identified international application was filed, and that, to the best of my knowledge and belief, the English translation of the international application No. PCT/FR00/02433 is a true and complete translation of the above identified international application as filed.

I hereby declare that all the statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the patent application issued thereon.

Date: 20 April 2001



Full name of the translator :

Elaine Patricia PARRISH

For and on behalf of RWS Group plc

Post Office Address :

Europa House, Marsham Way,
Gerrards Cross, Buckinghamshire,
England.

UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Jean Weissenbach, Jamilé Hazan

Application: CLONING, EXPRESSION AND CHARACTERIZATION OF THE
SPG4 GENE RESPONSIBLE FOR THE MOST COMMON FORM OF
AUTOSOMAL DOMINANT SPASTIC PARAPLEGIA

Serial No.: Herewith Art Unit:

Filing Date: Herewith Examiner:

Case: R-341894

PCT Application Information:

PCT Serial No: **PCT/FR00/02433** Priority Filing Date: September 3, 1999

PCT Filing Date: September 4, 2000

CERTIFICATE OF MAILING BY EXPRESS MAIL: "Express Mail" Mailing Label No. **EL713287403US**

I hereby certify that this paper and/or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated below and is addressed to the Commissioner for Patents, Washington, D.C. 20231

05/02/2001

Date of Deposit


Signature

Joseph Krieger

Typed/printed name of person signing

853 Sanders Road, #330
Northbrook, Illinois 60062
May 2, 2001

Box "PCT"
Commissioner for Patents
Washington, D.C. 20231

FIRST PRELIMINARY AMENDMENT

Sir:

Prior to the calculation of the application filing fees in connection with the

above-identified application, please amend the above-identified application as follows
(a marked up version of the amended claims are on the pages following the remarks):

IN THE SPECIFICATION:

Page 1, following the title, insert as a centered title --BACKGROUND
OF THE INVENTION--.

Page 1, prior to the first full paragraph beginning at line 5 and following
the inserted centered title "BACKGROUND OF THE
INVENTION", insert as a subtitle --1. Field of the Invention--.

Page 1, prior to the second full paragraph beginning at line 12, insert as
a subtitle --2. Background of the Invention--.

Page 2, prior to the paragraph beginning at line 32, insert as a centered
title --SUMMARY OF THE INVENTION--.

Page 21, line 30, replace paragraph "LEGENDS OF THE FIGURES"
with a centered title --BRIEF DESCRIPTION OF THE
DRAWINGS--.

Page 24, prior to line 1, insert as a centered title --DETAILED
DESCRIPTION OF THE PREFERRED EMBODIMENTS--.

Page 43, following the last paragraph, insert:

What is claimed and desired to be secured by Letters Patent of the United States is:--.

IN THE CLAIMS

Rewrite claims 3-5, 8-10, 12-19, 21, 23-26, and 29-36 as follows:

3. (amended) Purified or isolated nucleic acid according to claim 1, characterized in that it comprises at least one sequence of at least 15 consecutive nucleotides of the nt 714-809, ends inclusive, fragment of the sequence SEQ ID No. 2, of the sequence complementary thereto or of the sequence of the corresponding RNA thereof.

4. (amended) Purified or isolated nucleic acid according to claim 1, characterized in that it comprises a mutation corresponding to a natural polymorphism in humans.

5. (amended) Probe or primer, characterized in that it comprises a sequence of a nucleic acid according to claim 1.

8. (amended) Method for screening cDNA or genomic DNA libraries, or for cloning isolated genomic or cDNA encoding spastin, characterized in that it uses a nucleic acid sequence according to claim 1.

9. (amended) Method according to claim 8, for identifying the genomic or cDNA sequence of the SPG4 gene of mammals.
10. (amended) Method for identifying a mutation carried by the human SPG4 gene, characterized in that it uses a nucleic acid sequence according to claim 1.
12. (amended) Method for identifying the nucleic acid sequences which promote and/or regulate the expression of the SPG4 gene, characterized in that it uses a nucleic acid sequence according to claim 1.
13. (amended) Nucleic acid identified using a method according to claim 9.
14. (amended) Polypeptide encoded by a nucleic acid according to claim 1.
15. (amended) Polypeptide according to claim 14, with the exception of the 584 amino acid peptide, the sequence of which is identified in the GenBank databank under the accession number AB029006.
16. (amended) Polypeptide according to claim 14, characterized in that it comprises an amino acid sequence chosen from the group comprising:

9. (amended) Method according to claim 8, for identifying the genomic or cDNA sequence of the SPG4 gene of mammals.
10. (amended) Method for identifying a mutation carried by the human SPG4 gene, characterized in that it uses a nucleic acid sequence according to claim 1.
12. (amended) Method for identifying the nucleic acid sequences which promote and/or regulate the expression of the SPG4 gene, characterized in that it uses a nucleic acid sequence according to claim 1.
13. (amended) Nucleic acid identified using a method according to claim 9.
14. (amended) Polypeptide encoded by a nucleic acid according to claim 1.
15. (amended) Polypeptide according to claim 14, with the exception of the 584 amino acid peptide, the sequence of which is identified in the GenBank databank under the accession number AB029006.
16. (amended) Polypeptide according to claim 14, characterized in that it comprises an amino acid sequence chosen from the group comprising:

a) the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107 or the sequence of at least 10 consecutive amino acids of one of these sequences; and

b) the sequences which are homologs or variants of the sequences SEQ ID No. 3, SEQ ID No. 73 or SEQ ID No. 107.

17. (amended) Polypeptide according to claim 14, characterized in that it comprises the sequence of at least 8 consecutive amino acids of the sequence of the aa 197-228, ends inclusive, fragment of the sequence SEQ ID No. 3.

18. (amended) Polypeptide according to claim 14, characterized in that it comprises an amino acid sequence chosen from the group comprising the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107, which sequences carrying at least one of the mutations corresponding to a natural polymorphism in humans, and the sequences of the fragments thereof of at least 10 consecutive amino acids.

19. (amended) Cloning and/or expression vector containing a nucleic acid sequence according to claim 1.

21. (amended) Host cell transformed with a vector according to claim 19.

23. (amended) Mammal, except a human, according to claim 22, comprising a transformed cell, characterized in that the sequence of at least one of the two alleles of the SPG4 gene contains at least one of the mutations corresponding to a natural polymorphism in humans.

24. (amended) Use of a nucleic acid sequence according to claim 5, as a probe or primer, for detecting and/or amplifying nucleic acid sequences.

25. (amended) Use of a nucleic acid sequence according to claim 1, for screening a genomic or cDNA library.

26. (amended) Use of a nucleic acid sequence according to claim 1, for producing a recombinant or synthetic polypeptide.

29. (amended) Monoclonal or polyclonal antibodies or their fragments, chimeric antibodies or immunoconjugates, characterized in that they are capable of specifically recognizing a polypeptide according to claim 14.

30. (amended) Method for detecting and/or purifying a polypeptide, characterized in that it uses an antibody according to claim 29.

31. (amended) Method for genotypic diagnosis of AD-HSP associated with the SPG4 gene, characterized in that a nucleic acid sequence according to claim 1 is used.

32. (amended) Method for genotypic diagnosis of AD-HSP associated with the presence of at least one mutation on a sequence of the SPG4 gene, using a biological sample from a patient, characterized in that it includes the following steps:

a) where appropriate, isolation of the genomic DNA from the biological sample to be analyzed, or production of cDNA from the RNA of the biological sample;

b) specific amplification of said DNA sequence of the SPG4 gene likely to contain a mutation, using primers according to claim 5;

c) analysis of the amplification products obtained and comparison of their sequence with the corresponding normal sequence of the SPG4 gene.

33. (amended) Method for diagnosing AD-HSP associated with abnormal expression of a polypeptide encoded by the SPG4 gene, characterized in that one or more antibodies according to claim 29 is brought into contact with the biological material to be tested, under conditions which allow the possible formation of specific immunological complexes between said polypeptide and said antibody, and in that the immunological complexes possibly formed are detected and/or quantified.

34. (amended) Method for selecting a chemical or biochemical compound which is capable of modulating the expression or the activity of a polypeptide encoded by the SPG4 gene, characterized in that it comprises bringing a nucleic acid sequence according to claim 1 into contact with a candidate compound, and detecting a modification of the activity of said polypeptide.

35. (amended) Use of a nucleic acid sequence according to claim 1, for studying the expression or the activity of the SPG4 gene.

36. (amended) Kit for diagnosis, characterized in that it comprises at least a nucleic acid according to claim 5.

Add the following new claims 37-40:

37. (new claim) Method for selecting a chemical or biochemical compound which is capable of modulating the expression or the activity of a polypeptide encoded by the SPG4 gene, characterized in that it comprises bringing a nucleic acid sequence according to claim 14 into contact with a candidate compound, and detecting a modification of the activity of said polypeptide.

38. (new claim) Use of a polypeptide according to claim 14 for studying the expression or the activity of the SPG4 gene.

39. (new claim) Kit for diagnosis, characterized in that it comprises at least an antibody according to claim 29.

40. (new claim) Use of an antibody according to claim 29 for studying the expression or the activity of the SPG4 gene.

IN THE SEQUENCE LISTING

Line <110>, replace "CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE - CNRS" with the following:

-- Weissenbach, Jean
Hazan, Jamilé--.

Line <130>, replace "D18374" with --R-341894--.

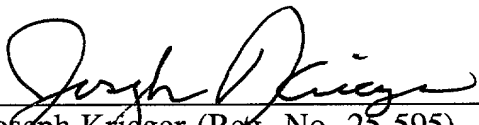
REMARKS

This Amendment is being presented in connection with the filing of the above-identified application which is a national phase application of the above-identified international (PCT) application. A substitute specification is being submitted concurrently with this Amendment and the above-identified application. This substitute specification is in accordance with the translation of the originally filed international application including the specification and claims; and an Abstract of the Disclosure (added as page 48 based on the abstract that was included in the published International Application (Publication No. WO 01/18198 A1)). In addition, seven (7) drawing sheets are being submitted concurrently with this Amendment and the above-identified application for use in connection with the above-identified application. These sheets are in accordance with the drawings appearing in International Publication No. WO 01/18198 A1.

With respect to the above amendments to the specification, subtitles have been added in order to conform the application to the requirements for applications of the United States Patent and Trademark Office. With respect to the above amendments to the claims, the claims have been amended principally so that each of the claims is dependent on a single claim rather than on multiple claims.

In accordance with 37 C.F.R. §§1.821-1.825, a sequence listing of 98 pages is being currently submitted herewith. This sequence listing is the listing submitted with the above-identified international application, but does incorporate the above amendments to line <110> (where the inventors were substituted for the assignee) and line <130> (where the attorney docket number for the above-identified national phase application was substituted for the attorney docket number of the international application). In further conformity with 37 C.F.R. §1.824, the sequence listing, as amended above, is being provided in computer readable form on a diskette in conformity with 37 C.F.R. §1.824(c)(1).

Respectfully submitted,
Mason, Kolehmainen, Rathburn & Wyss
(Customer #008668)

By 
Joseph Krieger (Reg. No. 25,595)
853 Sanders, #330
Northbrook, Illinois 60062
Telephone: 847-509-3720
Facsimile: 847-509-3722
e-mail: j_krieger@compuserve.com

Version of Amended Claim with Markings to Show Changes Made

The following is a marked up version of claims showing the amendments made to that claim (the changes are shown by underlining added matter and striking through deleted matter):

IN THE CLAIMS

Rewrite claims 3-5, 8-10, 12-19, 21, 23-26, and 29-36 as follows:

3. (amended) Purified or isolated nucleic acid according to claim 1-~~or~~ 2, characterized in that it comprises at least one sequence of at least 15 consecutive nucleotides of the nt 714-809, ends inclusive, fragment of the sequence SEQ ID No. 2, of the sequence complementary thereto or of the sequence of the corresponding RNA thereof.

4. (amended) Purified or isolated nucleic acid according to ~~one of~~ claim 1, characterized in that it comprises a mutation corresponding to a natural polymorphism in humans.

5. (amended) Probe or primer, characterized in that it comprises a sequence of a nucleic acid according to ~~one of claims~~ claim 1 ~~to 4~~.

8. (amended) Method for screening cDNA or genomic DNA libraries, or for cloning isolated genomic or cDNA encoding spastin, characterized in that it uses a nucleic acid sequence according to ~~one of claims~~claim 1 to 7.

9. (amended) Method according to claim 8, for identifying the genomic or cDNA sequence of the SPG4 gene of mammals, ~~in particular of mice~~.

10. (amended) Method for identifying a mutation carried by the human SPG4 gene, characterized in that it uses a nucleic acid sequence according to ~~one of~~claimsclaim 1 to 7.

12. (amended) Method for identifying the nucleic acid sequences which promote and/or regulate the expression of the SPG4 gene, characterized in that it uses a nucleic acid sequence according to ~~one of claims~~claim 1 to 7.

13. (amended) Nucleic acid identified using a method according to ~~one of claims~~claim 9 to 12.

14. (amended) Polypeptide encoded by a nucleic acid according to ~~one of claims~~claim 1 to 4 and 13.

15. (amended) Polypeptide according to claim 14, ~~preferably~~ with the exception of the 584 amino acid peptide, the sequence of which is identified in the GenBank databank under the accession number AB029006.

16. (amended) Polypeptide according to claim 14 ~~or 15~~, characterized in that it comprises an amino acid sequence chosen from the group comprising:

a) the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107 or the sequence of at least 10 consecutive amino acids of one of these sequences; and

b) the sequences which are homologs or variants of the sequences SEQ ID No. 3, SEQ ID No. 73 or SEQ ID No. 107.

17. (amended) Polypeptide according to claim 14 ~~or 15~~, characterized in that it comprises the sequence of at least 8 consecutive amino acids of the sequence of the aa 197-228, ends inclusive, fragment of the sequence SEQ ID No. 3.

18. (amended) Polypeptide according to claim 14 ~~or 15~~, characterized in that it comprises an amino acid sequence chosen from the group comprising the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107, which sequences carrying at least one of the mutations corresponding to a natural

polymorphism in humans, and the sequences of the fragments thereof of at least 10 consecutive amino acids.

19. (amended) Cloning and/or expression vector containing a nucleic acid sequence according to ~~one of claims~~claim 1 to 4, and 13.

21. (amended) Host cell transformed with a vector according to claim 19 or 20.

23. (amended) Mammal, except a human, according to claim 22, comprising a transformed cell, characterized in that the sequence of at least one of the two alleles of the SPG4 gene contains at least one of the mutations corresponding to a natural polymorphism in humans ~~or identified using a method according to claim 10 or 11.~~

24. (amended) Use of a nucleic acid sequence according to ~~one of~~ claimsclaim 5, ~~6 and 13,~~ as a probe or primer, for detecting and/or amplifying nucleic acid sequences.

25. (amended) Use of a nucleic acid sequence according to ~~one of~~ claimsclaim 1 to 7, and 13, for screening a genomic or cDNA library.

26. (amended) Use of a nucleic acid sequence according to ~~one of~~
~~claims~~claim 1 to 4 and 13, for producing a recombinant or synthetic polypeptide.

29. (amended) ~~One~~Monoclonal or polyclonal antibodies or their
fragments, chimeric antibodies or immunoconjugates, characterized in that they are
capable of specifically recognizing a polypeptide according to ~~one of claims~~claim 14 to
18, and 28.

30. (amended) Method for detecting and/or purifying a polypeptide
~~according to one of claims 14 to 18, and 28,~~ characterized in that it uses an antibody
according to claim 29.

31. (amended) Method for genotypic diagnosis of AD-HSP associated
with the SPG4 gene, characterized in that a nucleic acid sequence according to ~~one of~~
~~claims~~claim 1 to 7 and 13 is used.

32. (amended) Method for genotypic diagnosis of AD-HSP associated
with the presence of at least one mutation on a sequence of the SPG4 gene, using a
biological sample from a patient, characterized in that it includes the following steps:

a) where appropriate, isolation of the genomic DNA from the
biological sample to be analyzed, or production of cDNA from the RNA of the

biological sample;

b) specific amplification of said DNA sequence of the SPG4 gene likely to contain a mutation, using primers according to either of ~~claims~~claim 5 and 6 or a nucleic acid according to claim 13;

c) analysis of the amplification products obtained and comparison of their sequence with the corresponding normal sequence of the SPG4 gene.

33. (amended) Method for diagnosing AD-HSP associated with abnormal expression of a polypeptide encoded by the SPG4 gene, characterized in that one or more antibodies according to claim 29 is ~~(are)~~ brought into contact with the biological material to be tested, under conditions which allow the possible formation of specific immunological complexes between said polypeptide and said antibody ~~or antibodies~~, and in that the immunological complexes possibly formed are detected and/or quantified.

34. (amended) Method for selecting a chemical or biochemical compound which is capable of ~~interacting directly or indirectly with a polypeptide according to one of claims 14 to 18, and 28, or with a nucleic acid according to one of claims 1 to 7, and 13, and/or which makes it possible to modulate~~modulating the

expression or the activity of these ~~a polypeptides~~ polypeptide encoded by the SPG4 gene, characterized in that it comprises bringing a nucleic acid sequence according to one of claims claim 1 to 7, and 13, a polypeptide according to one of claims 14 to 18, and 28, a vector according to either of claims 19 and 20, a cell according to claim 21, a mammal according to either of claims 22 and 23 or an antibody according to claim 29 into contact with a candidate compound, and detecting a modification of the activity of said polypeptide.

35. (amended) Use of a nucleic acid sequence according to one of claims claim 1 to 7, and 13, of a polypeptide according to one of claims 14 to 18, and 28, of a vector according to either of claims 19 and 20, of a cell according to claim 21, of a mammal according to either of claims 22 and 23 or of an antibody according to claim 29, for studying the expression or the activity of the SPG4 gene.

36. (amended) Kit or pack for diagnosis, characterized in that it comprises at least one compound chosen from the following group of compounds:

————— a) ——— a nucleic acid according to either of claims claim 5 and 6;

and

————— b) ——— an antibody according to claim 29.

ATTORNEY DOCKET: R-341894

Substitute Specification
for

**CLONING, EXPRESSION AND CHARACTERIZATION
OF THE SPG4 GENE RESPONSIBLE FOR THE MOST
COMMON FORM OF AUTOSOMAL DOMINANT
SPASTIC PARAPLEGIA**

Inventors:

Jean Weissenbach, Jamilé Hazan

CERTIFICATE OF MAILING BY EXPRESS MAIL:

"Express Mail" Mailing Label No. **EL713287403US.**

I hereby certify that this paper and/or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated below and is addressed to the Commissioner for Patents, Washington, D.C. 20231


Signature

05/02/2001
Date of Deposit

Joseph Krieger
Typed/printed name of person signing

CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE
RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL DOMINANT
SPASTIC PARAPLEGIA.

5 The invention relates to the identification and characterization of the SPG4 gene
encoding spastin, which is responsible for the most common form of autosomal
dominant hereditary spastic paraplegia (HSP), to the cloning and characterization of its
cDNA, and also to the corresponding polypeptides. The invention also relates to
10 vectors, to transformed cells and to transgenic animals, and also to diagnostic methods
and kits and to methods for selecting a chemical or biochemical compound capable of
interacting directly or indirectly with a polypeptide according to the invention.

Hereditary spastic paraplegias (HSPs) are degenerative disorders of the central
nervous system, characterized by bilateral and progressive spasticity of the lower limbs.
They reveal themselves clinically through difficulties in walking possibly evolving into
15 total paralysis of both legs. The physiopathology of this set of diseases is, to date,
relatively undocumented; however, anatomopathological data make it possible to
conclude that the attack is limited to the pyramidal tracts responsible for voluntary
motricity in the spinal cord (Reid, 1997). Various clinical and genetic forms of HSP
exist. The so-called "pure" HSPs, which correspond to isolated spasticity of the lower
20 limbs, are clinically distinguished from the "complex" HSPs, for which the spasticity of
the legs is associated with other clinical signs of neurological or non-neurological type
(Bruyn et al., 1991). From a genetic point of view, the HSPs can be transmitted
according to the autosomal dominant (AD-HSP), autosomal recessive (AR-HSP) or X-
linked (X-HSP) mode. The "pure" form of HSP, which is most commonly transmitted
25 according to the autosomal dominant mode, remains the most frequent (approximately
80% of HSPs) (Reid, 1997). The incidence of HSPs, which remains difficult to estimate
because of rare epidemiological studies and the considerable clinical variability, varies
from 0.9 : 100 000 in Denmark, 3 to 9.6 : 100 000 in certain regions of Spain (Polo et
al., 1991) or 14 : 100 000 in Norway (Skre, 1974) (approximately 3 : 100 000 in
30 France).

In addition to this great clinical variability, which is observed not only between
various families but also between various affected members of the same family, the
HSPs are also characterized by considerable genetic heterogeneity. In the case of
AD-HSPs, four loci have been identified, to date, on chromosomes 14 (locus SPG3)
35 (Hazan et al., 1993), 2 (locus SPG4) (Hazan et al., 1994; Hentali et al., 1994), 15

(locus SPG6) (Fink et al., 1995) and 8 (locus SPG8) (Hedera et al., 1999). The study of a large number of families exhibiting an AD-HSP has shown that the gene carried by chromosome 2 is a main locus of this form of the disease, found in 40 to 50% of the families analyzed (The Hereditary Spastic Paraplegia Working Group, 1996; Durr et al., 1996). An anticipation phenomenon was observed in some locus SPG4-linked HSP families; this phenomenon has, subsequently, been associated with the expansion of a (CAG)_n repeat demonstrated in 6 Danish families (Nielsen et al., 1997) using the RED (for Rapid Expansion Detection) technique. It has, however, never been possible to confirm this expansion in any of the families tested by this method or by the systematic search for sequences of (CAG)_n type in physical maps composed of YAC (for Yeast Artificial Chromosome) or BAC (for Bacterial Artificial Chromosome) clones (Hazan et al., Genomics, 60 (3), 309-19, 1999).

To date, three genes responsible for two forms of X-HSP and one form of AR-HSP have been identified. Mutations in the gene which encodes a neuron-specific cell adhesion molecule, L1-CAM (for L1 Cell Adhesion Molecule), and which is located at Xq28 (locus SPG1) cause a complex form of HSP (Jouet et al., 1994) in which the spasticity is associated with a mental handicap, whereas mutations in the PLP (for ProteoLipid Protein) gene located at Xq21 (locus SPG2), which encodes a constitutive molecule of the myelin layer, cause pure and complex forms of X-HSP (Saugier-Veber, P. et al., 1994). More recently, mutations in the gene located at 16q24.3 (locus SPG7), which encodes paraplegin, a mitochondrial ATPase of the AAA (for "ATPases Associated with diverse cellular Activities") protein family (Confalonieri et al., 1995), have been associated with complex and pure forms of AR-HSP (Casari et al., 1998).

Thus, there remains, today, a great need to identify and characterize the gene responsible for the most common form of AD-HSP. The identification of this gene should, in particular, allow, besides the possibility of a test for antenatal screening in the families concerned, a better understanding of some of the molecular mechanisms engendering these degenerations specific for nerve bundles of the spinal cord, or even make it possible to provide an elementary response regarding therapeutic treatment for the patients.

This is precisely the subject of the present invention.

After having delimited the localization range between the D2S352 and D2S2347 genetic markers by studying recombination events in locus SPG4-linked HSP families, the inventors have established a contig of BACs covering a physical distance evaluated at approximately 1.5 Mb and have undertaken a positional cloning strategy based on

sequencing the SPG4 range in order to completely identify all the genes located in the candidate region. The analysis of the sequence of the two BACs, D (b336P14) and G (B763N4), has revealed the presence of a gene which is composed of 17 exons, extending over a distance of approximately 100 kb, and which exhibits homology with the genes encoding proteins of the AAA family. Comparison of the sequence of this gene between the healthy and affected individuals of AD-HSP families has made it possible to demonstrate various mutations in the patients.

A subject of the invention is thus the identification and characterization of the SPG4 (or SPAST) gene encoding a novel nuclear member of the AAA family, responsible for the most common form of AD-HSP.

In a first aspect, a subject of the present invention is a purified or isolated nucleic acid of the SPG4 gene, characterized in that it comprises at least 15 consecutive nucleotides, preferably 20, 25, 30, 35, 40, 45, 50, 75, 100 or 200 consecutive nucleotides, of a sequence chosen from the group comprising:

- the sequence SEQ ID No. 1, which is a genomic sequence of the human SPG4 gene;
- the nucleic acid sequences which are homologs or variants of the nucleic acid of sequence SEQ ID No. 1;
- the sequence which is complementary thereto; and
- the sequence of the corresponding RNA thereof.

The present invention relates, of course, to both the DNA and RNA sequences, and also the sequences which hybridize with them, as well as the corresponding double-stranded DNAs.

The terms "nucleic acid", "nucleic acid sequence" or "sequence of nucleic acid", "polynucleotide", "oligonucleotide", "polynucleotide sequence", and "nucleotide sequence", which will be used equally in the present description, will be intended to refer to both a double-stranded DNA, a single-stranded DNA and products of transcription of said DNAs, and/or an RNA fragment, said isolated natural, or synthetic fragments which may or may not include unnatural nucleotides, referring to a precise series of nucleotides, which may or may not be modified, making it possible to define a fragment or a region of a nucleic acid. The expression "natural isolated, or synthetic DNA and/or RNA fragment, which may or may not include unnatural nucleotides" is intended to mean a precise series of nucleotides, which may or may not be modified, making it possible to define a fragment, a segment or a region of a nucleic acid.

It should be understood that the present invention does not relate to the genomic nucleotide sequences in their natural chromosomal environment, i.e. in the

natural state. It involves sequences which have been isolated and/or purified, i.e. they have been removed directly or indirectly, for example by copying, their environment having been at least partially modified.

5 The term "homologous nucleic acid sequence" is intended to refer to the sequences which have, with respect to the reference nucleic acid sequence, certain modifications, such as in particular a deletion, a truncation, an extension, a chimeric fusion and/or a mutation, in particular a point mutation, and the nucleic acid sequence of which shows at least 80%, preferably 90% or 95%, identity after alignment, with the reference nucleic acid sequence.

10 For the purpose of the present invention, the term "percentage of identity" between two nucleic acid or amino acid sequences is intended to refer to a percentage of nucleotides or of amino acid residues which are identical between the two sequences to be compared, obtained after the best alignment, this percentage being purely statistical and the differences between the two sequences being distributed
15 randomly and throughout their length. Sequence comparisons between two nucleic acid or amino acid sequences are traditionally carried out by comparing these sequences after having optimally aligned them, said comparison being carried out by segment or by "window of comparison" in order to identify and compare local regions of sequence similarity. The optimal alignment of the sequences for comparison can be
20 produced, besides manually, by means of the local homology algorithm of Smith and Waterman (1981) [Ad. App. Math. 2:482], by means of the local homology algorithm of Needleman and Wunsch (1970) [J. Mol. Biol. 48:443], by means of the similarity search method of Pearson and Lipman (1988) [Proc. Natl. Acad. Sci. USA 85:2444], and by means of computer programs using these algorithms (GAP, BESTFIT, FASTA and
25 TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI, or with the BLAST N or BLAST P comparison programs).

The percentage of identity between two nucleic acid or amino acid sequences is determined by comparing these two optimally aligned sequences by window of comparison in which the region of the nucleic acid or amino acid sequence to be
30 compared can comprise additions or deletions with respect to the reference sequence for optimal alignment between these two sequences. The percentage of identity is calculated by determining the number of identical positions for which the nucleotide or the amino acid residue is identical between the two sequences, dividing this number of identical positions by the total number of positions in the window of comparison and

multiplying the result obtained by 100 so as to obtain the percentage of identity between these two sequences.

For example, the BLAST program "BLAST 2 sequences" (Tatusova et al., "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol. Lett. 174:247-250), available on the site <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>, may be used, the parameters used being those given by default (in particular for the parameters "open gap penalty" : 5, and "extension gap penalty" : 2; the matrix chosen being, for example, the "BLOSUM 62" matrix proposed by the program), the percentage of identity between the two sequences to be compared being calculated directly by the program.

It preferably involves sequences for which the complementary sequences are capable of hybridizing specifically with one of the sequences of the invention. Preferably, the specific or high stringency hybridization conditions will be such that they ensure at least 80%, preferably 90% or 95%, identity after alignment between one of the two sequences and the sequence which is complementary to the other.

Hybridization under high stringency conditions means that the temperature and ionic strength conditions are chosen such that they allow the hybridization between two complementary DNA fragments to be maintained. By way of illustration, high stringency conditions of the hybridization step for the purposes of defining the polynucleotide fragments described above are advantageously as follows.

The DNA-DNA or DNA-RNA hybridization is carried out in two steps: (1) prehybridization at 42°C for 3 hours in phosphate buffer (20 mM, pH 7.5) containing 5 x SSC (1 x SSC corresponds to a 0.15 M NaCl + 0.015 M sodium citrate solution), 50% of formamide, 7% of sodium dodecyl sulfate (SDS), 10 x Denhardt's, 5% of dextran sulfate and 1% of salmon sperm DNA; (2) actual hybridization for 20 hours at a temperature dependent on the size of the probe (i.e. 42°C for a probe of size > 100 nucleotides), followed by two 20-minute washes at 20°C in 2 x SSC + 2% SDS and one 20-minute wash at 20°C in 0.1 x SSC + 0.1% SDS. The final wash is carried out in 0.1 x SSC + 0.1% SDS for 30 minutes at 60°C for a probe of size > 100 nucleotides. The high stringency hybridization conditions described above for a polynucleotide of defined size will be adjusted by those skilled in the art for oligonucleotides of greater or smaller size, according to the teaching of Sambrook et al., 1989.

The term "nucleic acid sequence which is a variant" or "nucleic acid which is a variant" of a reference nucleic acid sequence will be intended to refer to the set of nucleic acid sequences corresponding to allelic variants, i.e. individual variations of the

reference nucleic acid sequence. These natural mutated sequences correspond to polymorphisms present in mammals, in particular in human beings, and in particular to polymorphisms which can cause a pathology to occur and/or to develop.

While the sequences according to the invention relate to normal sequences, 5 they also relate to sequences which are mutated insofar as they include at least one point mutation, and preferably at most 10% of mutations, with respect to the normal sequence.

In particular, the variant nucleic acid sequences will comprise any sequence of at least 15 consecutive nucleotides, preferably 20, 25, 30, 50, 100 or 200 consecutive nucleotides, of a polymorphic sequence of the genomic sequence of the human SPG4 gene of sequence SEQ ID No. 1, and the nucleic acid sequence of which has, with respect to the sequence SEQ ID No. 1, at least one mutation corresponding in particular to a truncation, deletion, substitution and/or addition of an amino acid residue. In the present case, the variant nucleic acid sequences having at least one mutation will herein be linked to the pathologies of AD-HSP type linked to SPG4 locus.

Preferably, the present invention relates to the mutated nucleic acid sequences in which the mutations produce a modification of the amino acid sequence of the polypeptide encoded by the normal sequence.

The term "variant nucleic acid sequences" will also be intended to refer to any
20 RNA or cDNA resulting from a mutation of a splice site of the genomic nucleic acid
sequence SEQ ID No. 1.

Preferably, the invention relates to a purified or isolated nucleic acid of the SPG4 gene according to the invention, characterized in that it comprises a sequence chosen from the group comprising:

- 25 a) the sequence SEQ ID No. 1, the sequence SEQ ID No. 2, the sequence SEQ ID No. 72, the sequence SEQ ID No. 106 or the sequence of at least 15, preferably 20, 25, 30, 35, 40, 45, 50, 75, 100 or 200, consecutive nucleotides of the sequence SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 or SEQ ID No. 106;
- 30 b) the nucleic acid sequences which are homologs or variants of the sequences SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 or SEQ ID No. 106; and
- c) the complementary sequence or the RNA sequence corresponding to the sequences as defined in a) and b),

preferably with the exception of the nucleic acid identified in the GenBank database under the accession number AB029006.

The nucleic acid the sequence of which is disclosed in the GenBank database under the accession number AB029006 corresponds to the sequence of one of the 100 cDNAs derived from a human brain mRNA library identified by the Kazusa DNA Research Institute in Japan (Kikuno et al., DNA Research, 6, 197-205, 1999).

5 Preferably, the invention relates to a purified or isolated nucleic acid according to the invention, characterized in that it comprises at least one sequence of at least 15 consecutive nucleotides, preferably 20, 25, 30, 50 or 75 consecutive nucleotides, of the nt 714-809, ends inclusive, fragment of the sequence SEQ ID No. 2, of the sequence complementary thereto or of the sequence of the corresponding RNA thereof.

10 The invention preferably relates to a purified or isolated nucleic acid according to the present invention, characterized in that it comprises a sequence chosen from the following group:

- the sequence SEQ ID No. 1;
- the sequence SEQ ID No. 2, which is the cDNA sequence encoding human spastin;
- 15 - the sequences SEQ ID No. 72 and SEQ ID No. 106, the sequence SEQ ID No. 72 representing the sequence of the incomplete cDNA encoding murine spastin represented in Figure 5, "mouse" line, and the SEQ ID No. 106 representing the complete sequence thereof;
- the nucleic acid sequences which are homologs or variants of the sequences SEQ ID
- 20 No. 1, SEQ ID No. 2, SEQ ID No. 72 or SEQ ID No. 106;
- the sequence complementary thereto; and
- the sequence of the corresponding RNA thereof.

25 Preferably, the invention relates to a purified or isolated nucleic acid according to the invention, characterized in that it comprises at least one mutation which corresponds to a natural polymorphism in humans, in particular the position and nature of which are identified in Table 5.

The primers or probes, characterized in that they comprise a sequence of a nucleic acid according to the invention, also form part of the invention.

30 The present invention thus relates to the set of primers which can be deduced from the nucleotide sequences of the invention and which may make it possible to demonstrate said nucleotide sequences of the invention, in particular the mutated sequences, using in particular an amplification method such as the PCR method, or a related method.

35 The present invention also relates to the set of probes which can be deduced from the nucleotide sequences of the invention, in particular from the sequences

capable of hybridizing with them, and which may make it possible to demonstrate said nucleotide sequences, in particular to distinguish the normal sequences from the mutated sequences.

5 The present invention relates, in particular, to the probes or primers having sequences chosen from the sequences SEQ ID No. 4 to SEQ ID No. 71.

The invention also relates to the use of a nucleic acid sequence according to the invention as a probe or primer, for detecting, identifying, assaying or amplifying a nucleic acid sequence.

10 According to the invention, the polynucleotides which can be used as a probe or as a primer in processes for detecting, identifying, assaying or amplifying a nucleic acid sequence will have a minimum size of 15 bases, preferably of 20 bases, or better still of 25 to 30 bases.

15 The set of probes and primers according to the invention may be labeled directly or indirectly with a radioactive or nonradioactive compound, using methods well known to those skilled in the art, in order to obtain a detectable and/or quantifiable signal.

The nonlabeled polynucleotide sequences according to the invention can be used directly as a probe or primer.

20 The sequences are generally labeled so as to obtain sequences which can be used for many applications. The labeling of the primers or of the probes according to the invention is carried out with radioactive elements or with nonradioactive molecules.

25 Among the radioactive isotopes used, mention may be made of ^{32}P , ^{33}P , ^{35}S , ^3H or ^{125}I . The nonradioactive entities are selected from ligands, such as biotin, avidin or streptavidin, dioxygenin, haptens, colorants and luminescent agents, such as radioluminescent, chemiluminescent, bioluminescent, fluorescent or phosphorescent agents.

30 The polynucleotides according to the invention can thus be used as a primer and/or probe in processes using, in particular, the PCR (polymerase chain reaction) technique (Erich, 1989; Innis et al., 1990, and Rolfs et al., 1991). This technique requires choosing pairs of oligonucleotide primers framing the fragment which must be amplified. Reference may, for example, be made to the technique described in American patent US No. 4,683,202. The amplified fragments can be identified, for example after agarose or polyacrylamide gel electrophoresis, or after a chromatographic technique such as gel filtration or ion exchange chromatography, and
35 then sequenced. The specificity of amplification can be controlled using, as a primer,

the nucleotide sequences of polynucleotides of the invention and, as a matrix, plasmids containing these sequences or the derived amplification products. The amplified nucleotide fragments can be used as reagents in hybridization reactions in order to demonstrate the presence, in a biological sample, of a target nucleic acid having a sequence complementary to that of said amplified nucleotide fragments.

The invention is also directed toward the nucleic acids which can be obtained by amplification using primers according to the invention.

Other techniques for amplifying the target nucleic acid can be advantageously employed as an alternative to PCR (PCR-like), using pairs of primers having nucleotide sequences according to the invention. The term "PCR-like" will be intended to refer to all methods using direct or indirect reproductions of nucleic acid sequences, or in which the labeling systems have been amplified. These techniques are, of course, known. In general, they involve amplifying the DNA with a polymerase; when the sample of origin is an RNA, it is advisable to perform reverse transcription beforehand. There are, currently, a great many processes which enable this amplification, such as for example the SDA (Strand Displacement Amplification) technique (Walker et al., 1992), the TAS (Transcription-based Amplification System) technique described by Kwok et al. in 1989, the 3SR (Self-Sustained Sequence Replication) technique described by Guatelli et al. in 1990, the NASBA (Nucleic Acid Sequence Based Amplification) technique described by Kievitis et al. in 1991, the TMA (Transcription Mediated Amplification) technique, the LCR (Ligase Chain Reaction) technique described by Landegren et al. in 1988 and improved by Barany et al. in 1991, which uses a heat-stable ligase, the RCR (Repair Chain Reaction) technique described by Segev in 1992, the CPR (Cycling Probe Reaction) technique described by Duck et al. in 1990, and the Q-beta-replicase amplification technique described by Miele et al. in 1983 and improved, in particular, by Chu et al. in 1986 and Lizardi et al. in 1988, and then by Burg et al., and also by Stone et al., in 1996.

When the target polynucleotide to be detected is an mRNA, use will advantageously be made, prior to carrying out an amplification reaction using the primers according to the invention or carrying out a detection process using the probes of the invention, of an enzyme of reverse transcriptase type in order to obtain a cDNA from the mRNA contained in the biological sample. The cDNA obtained will then serve as a target for the primers or probes used in the amplification or detection process according to the invention.

The probe hybridization technique can be carried out in diverse ways (Matthews et al., 1988). The most general method consists in immobilizing the nucleic acid extracted from the cells of various tissues or from cells in culture, on a support (such as nitrocellulose, nylon or polystyrene), and in incubating the immobilized target nucleic acid with the probe, under well defined conditions. After hybridization, the excess probe is eliminated and the hybrid molecules formed are detected using the appropriate method (measurement of the radioactivity, of the fluorescence or of the enzymatic activity linked to the probe).

According to another embodiment of the nucleic acid probes according to the invention, the latter can be used as a capture probe. In this case, a probe, termed "capture probe", is immobilized on a support and is used to capture, by specific hybridization, the target nucleic acid obtained from the biological sample to be tested, and the target nucleic acid is then detected using a second probe, termed "detection probe", labeled with an easily detectable element.

The splice acceptor or donor site sequences according to the present invention identified in Table 3 (sequences SEQ ID No. 74 to SEQ ID No. 105) also form part of the present invention.

In another aspect, the invention comprises a method for screening cDNA or genomic DNA libraries, or for cloning isolated genomic or cDNA encoding spastin, characterized in that it uses a nucleic acid sequence according to the invention.

Among these methods, mention may be made in particular of :

- the screening of cDNA libraries and the cloning of the isolated cDNAs (Sambrook et al., 1989; Suggs et al., 1981; Woo et al., 1979), using the nucleic acid sequences according to the invention;
- the screening of genomic libraries, for example of BACs (Chumakov et al., 1992; Chumakov et al., 1995), and, optionally, a genetic analysis by FISH (Cherif et al., 1990), using sequences according to the invention, enabling the isolation and chromosomal localization, and then the complete sequencing, of the SPG4 gene encoding spastin.

In particular, these methods according to the invention may be used for identifying and thus obtaining the genomic sequence or the cDNA of the SPG4 gene in other mammals, in particular mice.

These screening and/or cloning methods will comprise, in particular, a step of hybridization of a nucleic acid according to the invention with a nucleic acid contained in a genomic or cDNA library.

The invention also comprises a method for identifying the nucleic acid sequences which promote and/or regulate the expression of the SPG4 gene of sequence SEQ ID No. 1, characterized in that it uses a nucleic acid according to the invention.

5 The computer tools available to those skilled in the art enable them to easily identify, using the genomic nucleic acid sequences according to the invention, the promoter regulatory boxes required and sufficient for controlling gene expression, in particular the TATA, CCAAT and GC boxes, and also the stimulatory regulatory sequences ("enhancers"), or inhibitory regulatory sequences ("silencers"), which
10 control, in CIS, the expression of the genes according to the invention; among these regulatory sequences, mention should be made of IRE, MRE and CRE.

 The invention also relates to the methods for identifying mutations carried by the human SPG4 gene, in particular mutations responsible for autosomal dominant hereditary spastic paraplegia, characterized in that they use a nucleic acid sequence
15 according to the invention.

 These methods for identifying these mutations will, in particular, comprise the following steps: (i) isolation of the DNA from the biological sample to be analyzed, or production of a cDNA from the mRNA of the biological sample; (ii) specific amplification of the target DNA likely to have a mutation, using primers according to the invention;
20 (iii) analysis of the amplification products, in particular the size and/or the sequence of the amplification products, with respect to a reference sequence.

 The expression "methods for identifying a mutation according to the invention" is also intended to refer to a method which makes it possible to obtain the nucleic acid on which said mutation has been identified.

25 The promoter and/or regulatory sequences of the SPG4 gene according to the invention having mutations which may modify the expression of the corresponding protein also form part of the invention.

 The nucleic acids characterized in that they can be obtained using one of the preceding methods according to the invention, or the nucleic acids capable of
30 hybridizing, under high stringency conditions (homology of at least 80% between one of the two sequences and the sequence complementary to the other), with said nucleic acids, form part of the invention, especially the variant or homologous nucleic acids, in particular the nucleic acid sequences of allelic variants of the SPG4 gene of sequence SEQ ID No. 1 or of its cDNA of sequence SEQ ID No. 2, and also the genomic
35 sequences of the homologous genes of other mammals such as mice.

In the present description, the term "Spg4" will be intended to refer to the mouse gene homologous to the human SPG4 gene.

The use of a nucleic acid sequence according to the invention as a probe or primer for screening a genomic library or a cDNA of course forms part of the subject of the present invention.

In another aspect, the invention comprises a purified or isolated polypeptide encoded by a nucleic acid according to the invention, preferably with the exception of the 584 amino acid peptide, the sequence of which is identified in the GenBank database under the accession number AB029006.

In the present description, the term "polypeptide" will be used to refer equally to a protein or a peptide.

Preferably, the present invention relates to a polypeptide according to the invention, characterized in that it comprises an amino acid sequence chosen from the following group:

- the sequence SEQ ID No. 3, corresponding to human spastin encoded by the sequence SEQ ID No. 2 of the cDNA of the human SPG4 gene;
- the sequence SEQ ID No. 73, corresponding to a fragment of murine spastin encoded by the sequence SEQ ID No. 72 of the incomplete cDNA of the mouse Spg4 gene, the sequence SEQ ID No. 73 is represented in Figure 4A, "SPAST_MOUSE" line;
- the sequence SEQ ID No. 107, corresponding to murine spastin encoded by the sequence SEQ ID No. 106 of the complete cDNA of the mouse Spg4 gene;
- the sequences of polypeptides which are homologs and variants of the polypeptide of sequence SEQ ID No. 3, SEQ ID No. 73 or SEQ ID No. 107; and
- the sequences of the fragments thereof of at least 8, 10, 15, 30 or 50 consecutive amino acids.

Also preferably, a subject of the invention is a polypeptide according to the invention, characterized in that it comprises an amino acid sequence chosen from the group comprising:

- a) the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107 or the sequence of at least 10 consecutive amino acids of one of these sequences; and
- b) the sequences which are homologs or variants of the sequences SEQ ID No. 3, SEQ ID No. 73 or SEQ ID No. 107.

Also preferably, a subject of the invention is a polypeptide according to the invention, characterized in that it comprises the sequence of at least 8, preferably of at

least 10, 15, 20 or 30, consecutive amino acids of the sequence of the aa 197-228, ends inclusive, fragment of the sequence SEQ ID No. 3.

Also preferably, a subject of the invention is a polypeptide according to the invention, characterized in that it comprises an amino acid sequence chosen from the following group:

- the sequence SEQ ID No. 3, the sequence SEQ ID No. 73 and the sequence SEQ ID No. 107, which sequences carrying at least one of the mutations corresponding to a natural polymorphism in humans, in particular those the nature and location of which are identified in Table 5 hereinafter, or those which may be identified using the methods for identifying mutations of the SPG4 gene, according to the present invention; and
- the sequences of the fragments thereof of at least 8, 10, 15, 30 or 50 consecutive amino acids.

It should be understood that the invention does not relate to polypeptides in natural form, i.e. they are not taken in their environment. Specifically, the invention relates to the peptides which are obtained by purification from natural sources, or obtained by genetic recombination or by chemical synthesis, and which can therefore include unnatural amino acids. The production of a recombinant polypeptide, which can be carried out using one of the nucleotide sequences according to the invention, is particularly advantageous since it makes it possible to obtain an increased degree of purity of the desired polypeptide.

The term "homologous polypeptide" will be intended to refer to the polypeptides which have certain modifications with respect to the reference polypeptide, such as in particular one or more deletions or truncations, an extension, a chimeric fusion and/or one or more substitutions, and the amino acid sequence of which shows at least 80%, preferably 90% or 95%, identity after alignment, with the reference amino acid sequence.

The term "variant polypeptide" (or protein variant) will be intended to refer to the set of polypeptides encoded by the variant nucleic acid sequences as defined above.

In particular, the variant polypeptides will comprise any polypeptide which is encoded by the mutated genomic sequence of the SPG4 gene of sequence SEQ ID No. 1, and the amino acid sequence of which has at least one mutation corresponding in particular to a truncation, deletion, substitution and/or addition of amino acid residues with respect to the sequence SEQ ID No. 3. In the present case, the variant

polypeptides having at least one mutation will be linked to the pathologies of AD-HSP type.

The term "variant polypeptide" will also be intended to refer to any polypeptide resulting from mutation of a splice site in the genomic nucleic acid sequence SEQ ID No. 1.

The invention also comprises the cloning and/or expression vectors containing a nucleic acid sequence according to the invention.

The vectors according to the invention, characterized in that they include the elements which allow the expression and/or the secretion of said sequences in a host cell, or a cellular addressing sequence, also form part of the invention.

The vectors characterized in that they include a promoter and/or regulator sequence according to the invention also form part of the invention.

Said vectors will preferably include a promoter, translation initiation and termination signals, and also suitable regions for regulating the transcription. They should be able to be maintained stably in the cell and can, optionally, have particular signals which specify secretion of the translated protein.

These various control signals are chosen as a function of the host cell used. To this effect, the nucleic acid sequences according to the invention can be inserted into vectors which replicate autonomously in the host chosen, or vectors which integrate in the host chosen.

Among the systems which replicate autonomously, use will preferably be made, as a function of the host cell, of the systems of plasmid or viral type, the viral vectors possibly in particular being adenoviruses (Perricaudet et al., 1992), retroviruses, lentiviruses, poxviruses or herpesviruses (Epstein et al., 1992). Those skilled in the art know the technology which can be used for each of these systems.

When integration of the sequence into the chromosomes of the host cell is desired, use may be made, for example, of the systems of plasmid or viral type; such viruses will, for example, be retroviruses (Temin, 1986), or AAVs (Carter, 1993).

Among the nonviral vectors, preference is given to naked polynucleotides such as naked DNA or naked RNA according to the technique developed by the company VICAL, yeast artificial chromosomes (YAC) for expression in yeast, mouse artificial chromosomes (MAC) for expression in murine cells and, preferably, human artificial chromosomes (HAC) for expression in human cells.

Such vectors will be prepared according to the methods commonly used by those skilled in the art, and the clones resulting therefrom can be introduced into a

suitable host using standard methods, such as for example lipofection, electroporation or heat shock.

The invention also comprises the host cells, in particular the eukaryotic and prokaryotic cells, transformed with the vectors according to the invention, and also the
5 transgenic animals, except humans, comprising one of said transformed cells according to the invention.

Among the cells which can be used for these purposes, mention may of course be made of bacterial cells (Olins and Lee, 1993), but also yeast cells (Buckholz, 1993), as well as animal cells, in particular cultures of mammalian cells (Edwards and Aruffo,
10 1993), and especially Chinese hamster ovary (CHO) cells, but also insect cells in which it is possible to use processes implementing baculoviruses, for example (Luckow, 1993). A preferred cellular host for expressing the proteins of the invention consists of CHO cells.

Among the mammals according to the invention, preference will be given to
15 animals such as mice, rats or rabbits, expressing a polypeptide according to the invention.

Among the mammals according to the invention, preference will also be given to those comprising a transformed cell characterized in that the sequence of at least one of the two alleles of the SPG4 gene contains at least one of the mutations
20 corresponding to a natural polymorphism in humans, in particular those the nature and location of which are identified in Table 5 hereinafter, or those which may be identified using the methods for identifying a mutation of the SPG4 gene, according to the present invention.

Among the mammals according to the invention, preference will also be given to
25 animals such as mice, rats or rabbits, characterized in that the gene encoding spastin according to the invention is not functional or is knocked out.

Among the animal models more particularly advantageous herein, there are, in particular:

- the transgenic animals having, at least in one of their two allelic sequences of the
30 SPG4 gene, at least one of the mutations the position and nature of which are identified in Table 5 or identified using a method according to the present invention. These transgenic animals are obtained, for example, by homologous recombination on embryonic stem cells, transfer of these stem cells to embryos, selection of the chimeras affected in the reproductive lines, and growth of said chimeras;

- the transgenic animals (preferably mice) overexpressing the SPG4 gene into which one of said mutations according to the invention may be introduced. The mice are obtained, for example, by transfection of a copy of this gene under the control of a strong promoter which is ubiquitous in nature or selective for a tissue type, or after viral transcription;
- the transgenic animals (preferably mice) made deficient for the SPG4 gene according to the invention by inactivation using the LOXP/CRE recombinase system (Rohlmann et al., 1996) or any other system for inactivating the expression of this gene.

The cells and mammals according to the invention can be used in a method for producing a polypeptide according to the invention, as described below, and can also be used as a model for analysis and for DNA (genomic or cDNA) library screening.

The transformed cells or mammals as described above can thus be used as models in order to study the interactions between the polypeptides according to the invention, and chemical or protein compounds, which are involved directly or indirectly in the activities of the polypeptides according to the invention, this being in order to study the various mechanisms and interactions which come into play.

They can especially be used for selecting products which interact with the polypeptides according to the invention, in particular human spastin of sequence SEQ ID No. 3 or the variants thereof according to the invention, as a cofactor or as an inhibitor, in particular a competitive inhibitor, or which have agonist or antagonist activity for the activity of the polypeptides according to the invention. Preferably, said transformed cells or transgenic animals will be used as a model which, in particular, enables the selection of products which make it possible to combat the pathology linked to the SPG4 gene mentioned above.

The invention also relates to the use of a cell, of a mammal or of a polypeptide according to the invention for screening a chemical or biochemical compound which can interact directly or indirectly with the polypeptides according to the invention, and/or which is capable of modulating the expression or the activity of these polypeptides.

The invention also relates to the use of a nucleic acid sequence according to the invention for synthesizing recombinant polypeptides.

The method for producing a polypeptide of the invention in recombinant form is, itself, included in the present invention, and is characterized in that the transformed cells, in particular the cells or mammals of the present invention, are cultured under conditions which allow the expression of a recombinant polypeptide encoded by a

nucleic acid sequence according to the invention, and in that said recombinant polypeptide is recovered.

The recombinant polypeptides, characterized in that they can be obtained using said production method, also form part of the invention.

5 The recombinant polypeptides obtained as indicated above can be in both glycosylated and nonglycosylated form and may or may not have the natural tertiary structure.

10 These polypeptides can be produced based on the nucleic acid sequences defined above, according to the techniques for producing recombinant polypeptides known to those skilled in the art. In this case, the nucleic acid sequence used is placed under the control of signals which allow its expression in a cellular host.

An effective system for producing a recombinant polypeptide requires a vector and a host cell according to the invention.

15 These cells can be obtained by introducing into host cells a nucleotide sequences inserted into a vector as defined above, and then culturing said cells under conditions which allow the replication and/or expression of the transfected nucleotide sequence.

20 The processes for purifying a recombinant polypeptide which are used are known to those skilled in the art. The recombinant polypeptide can be purified from cell lysates and extracts and/or from the culture medium supernatant, with methods used individually or in combination, such as fractionation, chromatography methods, immunoaffinity techniques using specific monoclonal or polyclonal antibodies, etc.

25 The polypeptides according to the present invention can be obtained by chemical synthesis, this using one of the many known peptide syntheses, for example the techniques which implement solid phases or techniques which use partial solid phases, by condensation of fragments or by conventional synthesis in solution.

The solid-phase synthesis technique is well known to those skilled in the art. See in particular Stewart et al. (1984) and Bodansky (1984).

30 The polypeptides which are obtained by chemical synthesis and which can include corresponding unnatural amino acids are also included in the invention.

The mono- or polyclonal antibodies or their fragments, chimeric antibodies or immunoconjugates, characterized in that they are capable of specifically recognizing a polypeptide according to the invention, form part of the invention.

35 Specific polyclonal antibodies can be obtained from a serum of an animal immunized against the polypeptides according to the invention, in particular produced

by genetic recombination or by peptide synthesis, according to conventional procedures.

The advantage of antibodies which specifically recognize certain polypeptides, variants or immunogenic fragments thereof, according to the invention, will in particular
5 be noted.

The specific monoclonal antibodies can be obtained according to the conventional hybridoma culture method described by Köhler and Milstein, 1975.

The antibodies according to the invention are, for example, chimeric antibodies, humanized antibodies, or Fab or F(ab')₂ fragments. They can also be in the form of
10 labeled antibodies or immunoconjugates in order to obtain a detectable and/or quantifiable signal.

The invention also relates to methods for detecting and/or purifying a polypeptide according to the invention, characterized in that they use an antibody according to the invention.

The invention also comprises purified polypeptides, characterized in that they
15 are obtained using a method according to the invention.

Moreover, besides their use for purifying the polypeptides, the antibodies of the invention, in particular the monoclonal antibodies, can also be used for detecting these polypeptides in a biological sample.

They thus constitute a means of immunocytochemically or immuno-
20 histochemically analyzing the expression of the polypeptides according to the invention, in particular the polypeptide of sequence SEQ ID No. 3 or a variant thereof, on specific tissue sections, for example by immunofluorescence or gold labeling, or with an enzymatic immunoconjugates.

They may make it possible, in particular, to demonstrate abnormal expression
25 of these polypeptides in the biological samples or tissues, which makes them useful for monitoring the progression of the disease and the molecular diagnosis.

More generally, the antibodies of the invention can be advantageously used in any situation in which the expression of a normal or mutated polypeptide according to
30 the invention must be observed.

The methods for determining allelic variability, a mutation, a deletion, a loss of heterozygosity or any genetic abnormality of the SPG4 gene, according to the invention, characterized in that they use a nucleic acid sequence or an antibody according to the invention, also form part of the invention.

The present invention thus comprises a method for genotypic diagnosis of the pathology associated with the SPG4 gene, characterized in that a nucleic acid sequence according to the invention is used.

Preferably, the invention relates to a method for genotypic diagnosis of the disease associated with the presence of at least one mutation on a sequence of the SPG4 gene, using a biological sample from a patient, characterized in that it includes the following steps:

- a) where appropriate, isolation of the genomic DNA from the biological sample to be analyzed, or production of cDNA from the RNA of the biological sample;
- b) specific amplification of said DNA sequence of the SPG4 gene likely to contain a mutation, using primers according to the invention;
- c) analysis of the amplification products obtained and comparison of their sequence with the corresponding normal sequence of the SPG4 gene.

The invention also comprises a method for diagnosing the disease associated with abnormal expression of a polypeptide encoded by the SPG4 gene, in particular the polypeptide of sequence SEQ ID No. 3, characterized in that one or more antibodies according to the invention is (are) brought into contact with the biological material to be tested, under conditions which allow the possible formation of specific immunological complexes between said polypeptide and said antibody or antibodies, and in that the immunological complexes possibly formed are detected and/or quantified.

These methods are, for example, directed toward the methods for diagnosis, in particular antenatal diagnosis, of AD-HSP associated with the presence of a mutation in the SPG4 gene, according to the invention, by determining, using a biological sample from the patient, the presence of mutations in at least one of the sequences described above. The nucleic acid sequences analyzed may equally be genomic DNA, cDNA or mRNA.

Nucleic acids or antibodies based on the present invention may also be used to enable positive diagnosis in a patient or presymptomatic diagnosis in an individual at risk, in particular an individual with a family history of the disease.

There are, of course, a great number of methods which make it possible to demonstrate a mutation in a gene with respect to the wild-type gene. They can essentially be divided into two main categories. The first type of method is that in which the presence of a mutation is detected by comparing the mutated sequence with the corresponding wild-type sequence, and the second type is that in which the presence

of the mutation is detected indirectly, for example through evidence of mismatches due to the presence of the mutation.

These methods can use the probes and primers of the present invention which have been described. They are generally purified nucleic acid hybridization sequences comprising at least 15 nucleotides, preferably 20, 25 or 30 nucleotides, characterized in that they can hybridize specifically with a nucleic acid sequence according to the invention.

Preferably, the specific hybridization conditions are such as those defined above or in the examples. The length of these nucleic acid hybridization sequences can range from 15, 20 or 30 to 200 nucleotides, particularly from 20 to 50 nucleotides.

Among the methods for determining allelic variability, a mutation, a deletion, a loss of heterozygosity or a genetic abnormality, preference is given to the methods comprising at least one so-called PCR (polymerase chain reaction) or PCR-like amplification step for the target sequence according to the invention likely to have an abnormality, using a pair of primers having nucleotide sequences according to the invention. The amplified products may be treated with a suitable restriction enzyme before carrying out the detection and assaying of the product targeted.

The mutations of the SPG4 gene according to the invention may be responsible for various modifications of the translation product thereof, these modifications possibly being used for a diagnostic approach. Specifically, the antigenicity modifications linked to these mutations may allow the development of specific antibodies. The mutated gene product can be distinguished using these methods. All these modifications can be employed in a diagnostic approach, using several well-known methods based on the use of mono- or polyclonal antibodies which recognize the normal polypeptide or mutated variants, such as for example by RIA or by ELISA.

Thus, a subject of the invention is also a kit or pack for diagnosis, in particular for diagnosing AD-HSP associated with the presence of a mutation in the SPG4 gene, according to the invention, characterized in that it comprises at least one compound chosen from the following group of compounds:

- a) a nucleic acid, in particular as a primer or probe, according to the present invention; and
- b) an antibody according to the invention.

In another aspect, the invention comprises a method for selecting a chemical or biochemical compound capable of preventing and/or treating AD-HSP associated with the SPG4 gene, characterized in that a nucleic acid sequence according to the

invention, a polypeptide according to the invention, a vector according to the invention, a cell according to the invention, a mammal according to the invention or an antibody according to the invention is used.

The methods for selecting chemical or biochemical compounds capable of interacting directly or indirectly with polypeptides according to the invention or with the nucleic acids according to the invention, and/or making it possible to modulate the expression or the activity of these polypeptides, characterized in that they comprise bringing a polypeptide according to the invention, a transformed cell according to the invention or a mammal according to the invention into contact with a candidate compound, and detecting a modification of the activity of said polypeptide, are also included in the invention.

For example, but without being limited thereto, mention may be made of a method for identifying molecules capable of interacting with a polypeptide according to the invention, using a bacterial or yeast two hybrid system such as the Matchmaker Two Hybrid System 2, according to the instructions of the manual which is supplied with the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech).

The nucleic acids encoding proteins which interact with the promoter and/or regulatory sequences of the SPG4 gene, according to the invention, can be screened and/or selected using a one hybrid system such as that described in the manual which is supplied with the Matchmaker One Hybrid System kit from Clontech (Catalog No. K1603-).

In other aspect, the invention comprises the use of a nucleic acid or of a polypeptide according to the invention, of a vector according to the invention, of a cell according to the invention or of a mammal according to the invention, for studying the expression or the activity of the SPG4 gene.

Other characteristics and advantages of the invention appear in the remainder of the description with the examples and figures, the legends of which are given hereinafter.

LEGENDS OF THE FIGURES

FIGURES 1A, 1B and 1C : Physical map of the SPG4 range and genomic organization of SPG4.

FIGURE 1A : The 1.5 Mb candidate region is delimited by the D2S352 and D2S2347 genetic markers indicated in bold characters. The position of the polymorphic markers and other STSs is indicated in standard characters, whereas the position of

the ESTs is indicated in italics. The BAC clones constituting the presequencing map are represented by rectangles, with the name shown above and the precise size of the clone, if it could be determined, shown below. The name of the BACs A, B, C, etc. is followed by brackets containing the name of the clone preceded by a "b" if the clone is derived from the BACs library CITB_978_SKB, or by a "B" if it originates from the library RPCI-11.

FIGURE 1B : Schematic representation of the SPG4 gene which overlaps BACs D (b336P14) and G (B563N4). The exons are shown as black rectangles with their name above.

FIGURE 1C : The five mutations identified in seven SPG4 locus-linked AD-HSP families are positioned in exons 7, 11 and 13 and in the splice acceptor site of intron 15.

FIGURE 2 : Nucleic acid and protein sequence of the SPG4 cDNA of spastin.

The 17 vertical bars with a number located below represent the junctions between the various exons. The ATG initiator codon is located at nt position 126-128 and the STOP codon for termination is located at nt position 1974-1976. Five of the mutations identified to date, including the loss of exon 16, are indicated in italics (nt 1210, nt 1468, nt 1520, nt 1620 and for the loss of exon 16: nt 1813-1853). The polyadenylation site is in italics and underlined. The putative nuclear localization signal (NLS), RGKKK, and also the three conserved domains predicted by the analysis in the ProDom database are located at aa positions 7-11 (NLS), 342-409 (domain 92), 411-509 (domain 179) and 512-599 (domain 6226), respectively. The four motifs predicted by the sequence comparison in the Prosite database are: two "leucine zipper" motifs at aa positions 50-78 and 508-529, the ATP binding site (or Walker A motif) at aa positions 382-389 and the "helix-loop-helix" dimerization domain at aa positions 478-486. The Walker A and B motifs, "GPPGNGKT" and "IIFIDE", and also the AAA minimum consensus [lacuna] are underlined.

FIGURES 3A and 3B : Characterization of a splice site mutation in the affected individuals of three SPG4 locus-linked AD-HPS families.

FIGURE 3A : PCR amplification of fragment IV of the SPG4 cDNA using lymphoblast cDNA: well M, size marker VII (Boehringer); well 1, unaffected member of family 2992; well 2, patient of family 2992; well 3, unaffected member of family 5330; well 4, patient of family 5330; well 5, patient of family 5226; well 6, negative control (human genomic DNA).

FIGURE 3B : Sequence graph for the mutation of the splice acceptor site of intron 15.

Genomic sequence of the control individual above and of a patient of family 2992 below. The asterisk at nt position 1813-4 indicates an A->C polymorphism which affects a nonconserved nucleotide of the splice acceptor site of intron 15 in the patient.

FIGURES 4A and 4B : Spastin homologies.

The identical residues are highlighted by shaded areas.

FIGURE 4A : Multiple alignment created by CLUSTAL W of eight proteins derived from various organisms and having strong sequence homology with human spastin and murine spastin (SEQ ID No. 73).

FIGURE 4B : Alignment by CLUSTAL W of the yeast metalloproteases AFG3, RCA1 and YME1, and of human paraplegin and spastin.

FIGURE 5: Alignment by BLASTN of the nucleic acid sequences of the SPG4 cDNA and of its mouse ortholog Spg4 (SEQ ID No. 72). The polyadenylation site of the murine cDNA is underlined and in italics. The STOP codon is located at nt position 1515-1517 in the murine cDNA and at nt position 1974-1976 in the human cDNA.

FIGURES 6A, 6B and 6C : PCR analysis of the expression of SPG4 and of its murine ortholog Spg4.

FIGURE 6A : Collection of cDNA originating from multiple mouse tissues.

Well M, size marker V (Boehringer); well 1, heart; well 2, brain; well 3, spleen; well 4, lung; well 5, liver; well 6, skeletal muscle; well 7, kidney; well 8, testicle; well 9, E7 7-day embryo; well 10, E11 11-day embryo; well 11, E15 15-day embryo; well 12, E17 17-day embryo; well 13, negative control (mouse genomic DNA).

FIGURE 6B : Collection of cDNA originating from multiple human tissues.

Well M, size marker VII (Boehringer); well 1, brain; well 2, heart; well 3, kidney; well 4, liver; well 5, lung; well 6, pancreas; well 7, placenta; well 8, skeletal muscle; well 9, negative control (human genomic DNA); well 10, negative control (no DNA).

FIGURE 6C : Collection of cDNA originating from multiple human fetal tissues.

Well M, size marker VII (Boehringer); well 1, brain; well 2, heart; well 3, kidney; well 4, liver; well 5, lung; well 6, skeletal muscle; well 7, spleen; well 8, thymus; well 9, negative control (human genomic DNA); well 10, negative control (no DNA).

EXAMPLES

Example 1: Materials and methods

1) Subcloning and sequencing of the candidate region

Twelve BACs originating from two human genomic libraries, CITB_978_SKB (sold by Research Genetics) and RPCI-11 (Osoegawa et al., 1998), and covering the SPG4 range, were selected to be sequenced (Hazan et al., Genomics, 60 (3), 309-19, 1999). 40 µg of the DNA of each BAC were partially digested with the CviJI restriction enzyme (CHIMERx) and separated by electrophoresis on 0.4% LMP agarose gel (FMC). DNA fractions, the sizes of which vary in the region of 3, 5 and 10 kb, were eluted with β-agarase (Biolabs) and ligated to a plasmid vector pBAM3, which had been digested with SmaI and dephosphorylated, beforehand, in a ratio of 1 × insert per 5 × vector. Electrocompetent *E. coli* DH10B bacteria (GIBCO-BRL) were transformed with the various ligations, by electroporation. Approximately 1 000 to 1 500 subclones per BAC (8 to 10 equivalent genomes), consisting of 20% of clones with inserts at 10 kb, 40% of clones with inserts at 5 kb and 40% of clones with inserts at 3 kb, were isolated. The ends of the inserts of these clones were sequenced on a LICOR 4200 automatic sequencer. For each BAC, the sequences were assembled into a backbone consisting of several contigs, using the Phred and Phrap programs. The holes between each contig were sequenced with labeled dideoxynucleotides on an ABI 377 sequencer (PE-Applied Biosystems). The exons contained in these sequence contigs were predicted with the GRAIL II, GENSCAN, FGENEH and Genie computer programs. The sequences were also compared in the EMBL and GenBank nucleic acid and protein databases, with the BLASTN and BLASTX programs. The determination of the promoter sequences was carried out using the TSSG and TSSW computer programs. The results of all these sequence analyses were visualized using the Genotator sequence annotation program.

2) cDNA cloning

The cDNA of the SPG4 gene was isolated through 5' and 3' RACE-PCR experiments on polyA⁺ RNAs of fetal brain, adult brain and adult liver, using the Marathon cDNA amplification kit (Clontech) according to the supplier's instructions. A first PCR followed by an internal PCR were carried out with various pairs of primers, the sequences of which are indicated in Table 1 hereinafter:

Table 1
Primers used for the RACE-PCRs and the cDNA amplifications

Primer	Sequence (5'-3')	5' position pair/PCR product size			
SPA_5RACE5	CGGAGCTCCTCTTGGCTGCCATG (SEQ ID No.4)	nt 405			
SPA_5RACE6	AGAAGCGCTGGCAGAGCCACACGAAG (SEQ ID No.5)	nt 372			
SPA_5RACE7	AAGGCGACCAAACGCAGCAGCGCGAAG (SEQ ID No.6)	nt 331			
SPA_3RACE1	AGGAGCAAGCTGTGGAATGGTATAAG (SEQ ID No.7)	nt 550			
SPA_3RACE2	TGGTTATGGCCAAGGACCGCTTACAAC (SEQ ID No.8)	nt 689			
SPA_3RACE3	CAAACGGACGTCTATAATGACAGTAC (SEQ ID No.9)	nt 747			
SPA_3RACE4	TTAGGAATGTGGACAGCAACCTTGC (SEQ ID No.10)	nt 1075			
SPA_3RACE5	CTTCTCTGAGGCCTGAGTTGTTTAC (SEQ ID No.11)	nt 1207			
SPA_3RACE6	TGCTAGAATGACTGATGGATACTCAGG (SEQ ID No.12)	nt 1736			
SPA_3RACE7	AGATGCAGCACTGGGTCCTATCCG (SEQ ID No.13)	nt 1787			
SPA_3RACE8	ATGAACGTCATCGGCTACAGAAACAG (SEQ ID No.14)	nt 2037			
SPA_Db	TAGCAGTGGCTGCCGCCGT (SEQ ID No.15)	nt 45	b+m	655 bp	
SPA_Dm	AAGCGGTCCTTGGCCATAAC (SEQ ID No.16)	nt 700			
SPA_Dc	GGCGGCAGTGAGAGCTGTG (SEQ ID No.17)	nt 106	c+n	543 bp	
SPA_Dn	CTAGCTCTTTCACACTGTTC (SEQ ID No.18)	nt 649			
SPA_Ad	AACAGGCCTTCGAGTACATC (SEQ ID No.19)	nt 487	d+n	746 bp	
SPA_Am	CTGTGAACAACCTCAGGCCTC (SEQ ID No.20)	nt 1233			
SPA_Ac	ATGAGAAAGCAGGACAGAAG (SEQ ID No.21)	nt 532			
SPA_An	TGCCAAGTCTTGACCAGC (SEQ ID No.22)	nt 1175			
SPA_Ba	CTACAACTGCTACTCGTAAG (SEQ ID No.23)	nt 1036	a+m	763 bp	
SPA_Bm	CAGTGCTGCATCTTTTGCC (SEQ ID No.24)	nt 1799			
SPA_Bb	TAGGAATGTGGACAGCAACC (SEQ ID No.25)	nt 1076			
SPA_Bn	AAAGCTGTTAGGTCACTTCC (SEQ ID No.26)	nt 1780			
SPA_Ca	TGGAGATGACAGAGTACTTG (SEQ ID No.27)	nt 1550	a+m	766 bp	
SPA_Cm	CTGGAATACTTTCATCTGC (SEQ ID No.28)	nt 2316			
SPA_Cb	ATGAGGCTGTTCTCAGGCG (SEQ ID No.29)	nt 1603			

The RACE-PCR products were cloned with the TA-cloning kit (Invitrogen) and the corresponding clones were sequenced on an ABI 377 (PE-Applied Biosystems). The sequence of the SPG4 transcript was verified by sequencing PCR products amplified from a cDNA population originating from the lymphoblasts of 6 healthy individuals.

3) Detection of mutations

The total RNAs were extracted from lymphoblast lines of one affected individual per family studied and of 6 control individuals, using the RNA PLUS kit (bioprobe System). The cDNA synthesis was carried out on 500 ng to 1 µg of RNA, with 100 pmol of random hexameric primers (Pharmacia) and 200 units of Superscript II reverse transcriptase (Gibco BRL), under standard conditions. Four PCR amplifications, generating overlapping fragments which cover all of the SPG4 open reading frame, were carried out on the cDNAs of the patients and controls. Fragment I was amplified with the SPA_Db/SPA_Dm primers, and then by internal PCR with the SPA_Dc/SPA_Dn primers. Fragments II, III, and IV were amplified with the SPA_Ad/SPA_Am, SPA_Ba/SPA_Bm and SPA_Ca/SPA_Cm primers (cf. the sequences of these primers in Table 1), respectively. Each amplification was carried out in a total volume of 50 µl containing 4 µl of cDNA (~ 1/7th of the prep.), 20 pmol of each primer, 200 µM of dNTPs, 50 mM of KCl, 10 mM of Tris, pH 9, 1.5 mM MgCl₂, 0.1% of triton X-100, 0.01% of gelatin and 2.5 units of Taq polymerase (Cetus-PE). The PCR reactions were carried out according to the "hot start" process: the Taq polymerase is added at 92°C, after a first denaturation step of 5 min at 94°C. The samples are subsequently subjected to 35 cycles of denaturation (94°C for 40 sec), of hybridization (55°C for 50 sec, with the exception of fragment I: 58°C for 50 sec) and of elongation (72°C for 1 min), followed by a final elongation step (5 min at 72°C). The PCR products are sequenced on an ABI 377 automatic sequencer (PE-Applied Biosystems), with the SPA_Dc/SPA_Dn, SPA_Ac/SPA_An, SPA_Bb/SPA_Bn and SPA_Cb/SPA_Cm primers for fragments I, II, III and IV, respectively.

The mutations were also sought or confirmed by sequencing the 17 predicted exons of the SPG4 gene in the patients and controls. Each exon was amplified with the corresponding "a+m" pair of primers (cf. Table 2 hereinafter), with the exception of exon 1 (gSPAex1c/gSPAex1m), and exons 10, 11 and 12 which were co-amplified with the gSPAex10a/gSPAex12m and gSPAex11a/gSPAex12m pairs of primers.

Table 2
PCR primers for amplifying and sequencing the exons

Exon	Product size	PCR program	Primer	Sequence (5'-3') (SEQ ID Nos.; 30 to 71)
1	1048 bp	0	gSPAex1c	GTGAGCCGAACTGCACATTG
			gSPAex1m	CAAAGTCGACAGCTACAGTGC
			gSPAex1d	GGAAGTGTAGTTGAGTGGGA
			gSPAex1n	AGATGAGGCTCCGACCTAC
2	624 bp	3	gSPAex2a	AATGCCACACTTGTAACTCTC
			gSPAex2m	TGTGAATATATCATAATTTGGG
			gSPAex2b	TACAGCAGTTCTCATGATG
3	812 bp	1	gSPAex3a	GACCAAATTGGTGCATGCATG
			gSPAex3m	ACATTTCCAATACATCCCAC
4	379 bp	3	gSPAex4a	ATTTGTCAATTCACATGCAC
			gSPAex4m	TTAGAATGACTATACCTGAC
			gSPAex4n	TCAGGTAAAGTAAGACTC
5	830 bp	4	gSPAex5a	TTCCTATCTACCTAGTGAC
			gSPAex5m	TTTTATAGCAAGTTGCCCTG
			gSPAex5b	CCTATGAAGATCCTGGTAC
6	484 bp	3	gSPAex6a	TGTCATGATTCTAACAAGGG
			gSPAex6m	TCTATTTCACTCCTGACATG
7	420 bp	2	gSPAex7a	GTCATAGGGCTTAGGCTTC
			gSPAex7m	ATCATACTACCCACTTTTCC
8	647 bp	3	gSPAex8a	TGTTTGGGAAGATGCTACTG
			gSPAex8m	CTACTGAAGATAACGTACATG
9	1268 bp	1	gSPAex9a	CATTGATTGCCATGTATTGG
			gSPAex9m	AGAAGGCCAGAAATACTCAG
			gSPAex9b	GTACTTAAATCGGTAAATATGG
10]	1061 bp	4	gSPAex10a	CTCAAGTCTTAGGAATGCAG
11			gSPAex10b	GCACTTAACCAGGCTGTATG
12]	551 bp	3	gSPAex11a	CTCAGATGACTCACATAGC
			gSPAex12m	CTTTACTAGACTAATTCTCCTG

13	1361 bp	4	gSPAex13a	CAGATTCAAGAAGACAGATC
			gSPAex13m	GCAATAATTCACCACACTTG
			gSPAex13n	GGTAGTTCTTGTTTCTGCTC
14	985 bp	4	gSPAex14a	CAAGTGTGGTGAATTATTGC
			gSPAex14m	GAGCTGAAAAGTATTCAGC
			gSPAex14n	TGCAAAGGACATAGCCAGTG
15	1076 bp	1	gSPAex15a	AGCCTCTGGAGATAGTATGC
			gSPAex15m	CTAGAACAGGGGTCACAGTC
			gSPAex15n	TTGGACTTCTTAAACTTC
16	1404 bp	4	gSPAex16a	GCAGTATGCAAGAAATTGAAC
			gSPAex16m	GGCCTGTAATTTTCTTCTG
			gSPAex16b	GTACTGAATAGATACATGTAG
17	445 bp	3	gSPAex17a	GTGTAGCAGATCAACATAG
			gSPAex17m	CATCTTCAAGTTTGGTGCAC

Other than for exon 1, which is amplified using the Advantage GC genomic PCR kit (Clontech) according to the supplier's instructions, four slightly different PCR programs (1, 2, 3 and 4) were used to amplify the SPG4 exons (see Table 2). The amplifications were all carried out in a volume of 50 µl containing 100 ng of genomic DNA, 50 pmol of each primer, 250 µM pf dNTPs, 1X Takara buffer and 1 unit of Takara La Taq Taq polymerase (Shuzo Co.). The PCR reactions were carried out according to the "hot start" process: the Taq polymerase is added at 94°C, after a first denaturation step of 5 min at 96°C. The samples are subsequently subjected to 30 cycles of denaturation (94°C for 40 sec), of hybridization (prog. 1: 60°C for 50 sec; prog. 2: 58°C for 50 sec, prog. 3 and 4: 55°C for 50 sec) and of elongation (prog. 1 and 4: 72°C for 1 min, prog. 2 and 3: 72°C for 40 sec), followed by a final elongation step (10 min at 72°C). The sequencing of these PCR products was carried out on an ABI 377 sequencer (PE-Applied Biosystems), using either the PCR primers or the internal primers termed "b" and "n" (see Table 2).

4) Characterization of SPG4

The cDNA clones 977312 (EST AA560327) and 568234 (EST AA107866) derived from the mouse blastocyst and E8 embryo cDNA libraries, which both correspond to the murine ortholog of SPG4, were isolated using the IMAGE consortium

and sequenced in the laboratory on an ABI 377 sequencer (PE-Applied Biosystems). In order to analyze the expression profile of SPG4 and of its murine ortholog Spg4, the collections of cDNA from various fetal and adult human tissues, and also from mouse tissues (MTC panels, Clontech), were tested by PCR according to the supplier's protocol, with the SPA_Ca/SPA_Cm pair of primers for the human cDNAs and the SPA_Ca/spam (spam: 5'-ACCGAAGTCAAGAGCCTATC-3') pair for the mouse cDNAs. The PCR conditions are those used for amplifying SPG4 from lymphoblast line cDNA (cf. § Detection of mutations), except that these samples were subjected to 32 cycles for the cDNAs derived from adult human tissues and from mouse tissues, and to 28 cycles for the cDNAs derived from fetal tissues. The amplification products migrated by electrophoresis on 2% agarose gels.

5) Histological analysis of a muscle biopsy from a patient

The histological and histo-enzymatic analyses were carried out on a muscle biopsy from a patient derived from an SPG4 locus-linked family according to the standard techniques described in Casari et al., 1998.

6) Accession numbers in the public databases

The SPG4 (or SPAST) cDNA and the deduced protein sequence, GenBank/EMBL AJ246001; the incomplete Spg4 cDNA clone, GenBank/EMBL AJ246002; the SPG4 (or SPAST) gene, GenBank/EMBL AJ246003.

Example 2 : Analysis of the sequence of the SPG4 range

The analysis of the recombination events made it possible to reduce the SPG4 candidate region to a genetic range of 0 cM between the D2S352 and D2S2347 markers (19, 20). A presequencing map of the SPG4 range composed of 37 BACs was constructed (Hazan et al., in press in Genomics); the candidate region covers a physical distance of approximately of 1.5 Mb. Twelve overlapping BACs, stretching over the SPG4 region, with the exception of a single 4 kb hole between clones A and E, were selected to be sequenced (fig. 1A). Seven of these BACs (A, B, C, D, E, F and G), covering approximately 70% of the region of interest, have already been sequenced. The sequences of these 7 BACs were compared with those of the nucleic acid and protein databases, and analyzed with four exon prediction programs. These preliminary sequence analyses made it possible to reveal 14 potential transcription units, including three corresponding to the genes encoding xanthine dehydrogenase, steroid 5 α -reductase 2 and a TGF β -binding protein. Of the 14 genes detected by the sequence analysis, 9 had been previously identified in the EST (for "Expressed Sequence Tag") databases and located in the SPG4 range (Hazan et al., in press in

Genomics); the 5 remaining genes could only be identified by sequencing the candidate region. One of these 5 novel genes showed homology in 3' of its coding region, with the genes encoding the AAA protein family (Confalonieri et al., 1995). More thorough sequence analyses showed that this gene, named SPG4 (or SPAST), was composed of 17 exons and extended over a region of approximately 90 kb, covered by two adjacent BAC clones, D and G (cf. fig. 1B). The first three predicted exons of this gene were identified in BAC D, by two of the four exon prediction programs used, GRAIL II and GENSCAN; they show strong homology with a mouse blastocyst EST, AA560327. The last 14 exons are found in BAC G. The protein sequence deduced from exons 7 to 17 is significantly homologous to a subclass of the AAA family, which includes the Yta6p (Schnall et al., 1994), TBP6 (Schnall et al., 1994) and End 13 yeast proteins, and also the SKD1 mouse protein (Perier et al., 1994).

Of the four exon prediction programs FGENEH appears to be the most reliable and the most powerful, enabling detection of most of the genes of this chromosomal region at 2p21-p22. This observation also applies to the SPG4 gene, for which 15 exons could be demonstrated using this program, while only 4, 9 or 11 exons could be located using the Genie, GRAIL II and GENSCAN programs, respectively. The genomic organization of this gene (fig. 1B) could subsequently be confirmed by determining the sequence of the SPG4 cDNA. The intron/exon junctions are represented on table 3 hereinafter: the exon size ranges from 41 bp (exon 16) to 1.410 kb (exon 17), that of the introns ranging from 140 bp (intron 11) to 23.247 kb (intron 1).

Table 3

Intron/exon organization of the SPG4 gene

Exon/ intron	Exon size (bp)	Position on the cDNA	Splice acceptor site (SEQ ID No. 74 to 89)	Splice donor site (SEQ ID Nos. 90 to 105)	Intron size (bp)
1	540	1		TGAGAAAG/gtaactaggggctgg	23 247
2	87	541	atfttttattttaaag/CAGGACAG	AGGACAAAG/gtaagattgtatttgt	1 943
3	84	628	aatftttttctttcag/GTGAACAG	ACTTCTAG/gtatcaattaatgtat	9 190
4	96	712	cttctctgttgcatag/AGAAAGATG	CCAGTCAG/gtgggttaggttaac	15 745
5	188	808	actttttcctgtcag/AAAGTGGA	CTCATAAG/gtattctgggacagta	876
6	134	996	ttttgtatcctttaag/GGTACTCC	GTGGACAA/gtaagtttggccaict	283
7	94	1 130	aggctctgtttcttag/TGGAACAG	GGCCTGAG/gtaagaactttatatt	10 735
8	75	1 224	agtatatatttttag/TTGTTTAC	CAATGCTG/gtaagggttctctca	1 385
9	72	1 299	ctgtgatttttaaag/GCTAAAGC	CAAAATAC/gtgagtgctctgttcc	8 083
10	76	1 371	taatgctttgttttag/GTGGGAGA	TTTTATAG/gtaagaacatatattc	238
11	92	1 447	ctgtatttctctctag/ATGAAGTT	TTGATGGT/gtaagtggtgattatg	140
12	80	1 539	gatttttgcctgtag/GTACAGTC	GTTCTCAG/gtagggagattttat	4 715
13	43	1 619	ggatttttttttag/GCGTTTCA	ATGAGGAG/gtatgtatctgtgttt	1 389
14	80	1 662	ttttaaatattttcag/ACAAGACT	CTTGCTAG/gtgagtaatttggatt	1 521
15	71	1 742	tccttccctctctcag/AATGACTG	TATCCGAG/gtaggtatacaagagc	2 210
16	41	1 813	cttttatgtttttacag/AACTAAAA	CCAGTGAG/gtatagattttttaa	7 115
17	1 410	1 854	ctttttaaaaatctag/ATGAGAAA		

The sequences of the exons and introns are indicated in upper case and lower case, respectively.

Example 3 : Identification of the SPG4 cDNA

Several successive amplifications by 5' and 3' RACE-PCR were carried out on collections of adult liver and brain and fetal brain cDNA, in order to characterize the SPG4 transcript. All the 5' RACE-PCRs gave amplification products terminating at nt position 263 of the SPG4 cDNA (fig. 2), which was probably due to the rich GC content of the 5' region of the transcript (90% of GC in the 60 bp preceding nt position 263). Four overlapping PCR products, covering all of the coding region, were amplified from the cDNAs derived from the lymphoblasts of six control individuals, and entirely sequenced with the aim of verifying the sequence of the SPG4 transcript. Aligning the sequences of all the PCR and RACE-PCR products made it possible to reconstitute a 3263 bp sequence comprising a 1848 bp open reading frame preceded by a 125 bp untranslated 5' region (5' UTR for "5' UnTranslated Region") and followed by 1290 bp 3' UTR region including a polyadenylation site between nt positions 3227-3232, ~ 35 bp upstream of the polyA tail (fig. 2). Comparing the sequence of the SPG4 cDNA with the EST databanks made it possible to detect significant homology with 6 human ESTs, including EST N47973 which contains a more extended 3' noncoding region (+ 180 bp) comprising a second polyadenylation site. The translation initiation site was identified by the presence of a Kosak consensus sequence (CTGTGAatgA) defined as a "suitable context" for translation initiation given that a purine is located 3 nt upstream of the initiator ATG, itself preceded by a STOP codon. The 3263 bp cDNA sequence is identical to the transcribed sequence deduced from the 17 exons of the SPG4 gene. The analysis of the sequence of the 5' region using the TSSG and TSSW computer programs suggests the presence of a promoter sequence of the TATA box type located 43 bp upstream of nt position 1 of exon 1.

Example 4 : Mutations in the SPG4 gene

Heterozygous mutations were sought in the SPG4 cDNA originating from lymphoblasts of 14 patients derived from SPG4 locus-linked families (1 affected individual per family). Four overlapping PCR fragments, I, II, III and IV, covering the open reading frame of the SPG4 cDNA, were amplified and sequenced in the 14 patients, and also in 30 healthy control individuals. The agarose gel electrophoresis of PCR fragment IV showed three bands of equal intensity in 3 patients from families 2992, 5226 and 5330 originating from the same region of Switzerland, which would suggest a microdeletion or a mutation of a splice site; the two additional bands were not present in 2 healthy individuals derived from families 2992 and 5330 (fig. 3A). The genomic sequence of exon 16 revealed a heterozygous A->G mutation of the splice acceptor site (AG) of intron 15 in the affected

individuals of these three families (fig. 3B); this mutation engenders the loss of exon 16, followed by a reading frame shift in the abnormal transcript. None of the healthy members, including husbands and wives, carry this mutation of the splice site. The identification of the same mutation in all the affected members of these three Swiss families demonstrates the existence of a common ancestor, which had probably been suggested by the study of the haplotypes.

Three point mutations, 1210C->G, 1468G->A and 1620C->T, which introduced amino acid substitutions into the protein sequence (S362C, C448Y and R499C), were respectively revealed by sequencing PCR fragments III and IV in the affected individuals of families 624, 4014 and 618. These three substitutions all involve a cysteine residue, inducing the loss or insertion of a cysteine in the protein sequence. A 1 bp deletion, 1520delT, which creates the appearance of a STOP codon inducing a truncated protein composed of 465 amino acids (aa), was detected in the affected individuals of family A. None of the five mutations summarized in table 4 hereinafter was found in the control individuals tested, whether they belong to the healthy siblings or to the spouses of the seven families analyzed herein. These five mutations significantly affect the protein sequence in a very conserved domain, or AAA cassette (Beyer, 1997), which is composed of several protein motifs presumed to be responsible for the ATPase activity in all the members of the AAA family.

Table 4
Mutations in SPG4 in the patients suffering from AD-HSP

Family	Location	Mutation ^a	Amino acid change ^b	Consequence
624	exon 7	1 210 C → G	S362C	missense
4 014	exon 11	1 468 G → A	C448Y	missense
A	exon 11	1 520 delT	466STOPcodon	nonsense
618	exon 13	1 620 C → T	R499C	missense
2 992	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5 226	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5 330	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift

^a The nt positions refer to the sequence of the SPG4 cDNA.

^b The aa positions refer to the spastin sequence.

The bases of the exons are indicated in upper case, those of the introns in lower case.

PTC+7 aa = "premature termination codon" at 7 aa downstream of exon 16.

In addition to these five mutations described above, searches for heterozygous mutations, carried out on patients suffering from AD-HSP derived from 36 other families, made it possible to reveal 34 other mutations which modified or were likely to modify the product of expression of the SPG4 gene.

- 5 The characteristics of these 34 other mutations are summarized in table 5 hereinafter, into which the first five mutations mentioned above have also been inserted.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Table 5
Mutations in SPG4 in the patients suffering from AD-HSP

Family	Location	Mutation ^a	Amino acid change ^b	Consequence
624	exon 7	1210 C → G	S362C	missense
6958	exon 8	1233 G → A	G370R	missense
214	exon 8	1267 T → G	F381C	missense
1002	exon 8	1283 T → G	N386K	missense
027	exon 8	1288 A → G	K388R	missense
019	exon 10	1401 C → G	L426V	missense
4014	exon 11	1468 G → A	C448Y	missense
148	exon 11	1504 G → T	R460L	missense
618	exon 13	1620 C → T	R499C	missense
636	exon 15	1788 G → A	D555N	missense
627	exon 15	1792 C → T	A556V	missense
2971	exon 3	702 C → T	Q193STOP	nonsense
3655	exon 5	873 A → T	K229STOP	nonsense
1010	exon 5	907 C → A	S261STOP	nonsense
3938	exon 5	932 C → G	Y269STOP	nonsense
6922	exon 10	1416 C → T	R431STOP	nonsense
616	exon 10	1416 C → T	R431STOP	nonsense
605	exon 15	1809 C → T	R562STOP	nonsense
030	exon 2	578-579insA	PTC + 2 aa	shift + nonsense
615	exon 5	852del11	PTC + 18 aa	shift + nonsense
042	exon 5	882-883insA	PTC + 12 aa	shift + nonsense
032	exon 5	906delT	PTC + 17 aa	shift + nonsense
189	exon 9	1299delG	PTC + 3 aa	shift + nonsense
3686	exon 9	1340del5	PTC + 35 aa	shift + nonsense
625	exon 9	1340del5	PTC + 35 aa	shift + nonsense
A	exon 11	1520delT	PTC + 7 aa	shift + nonsense
115	exon 12	1574delGG	PTC + 2 aa	shift + nonsense
3266	exon 13	1634del22	PTC + 18 aa	shift + nonsense
149	exon 14	1684-1685insTT	PTC + 9 aa	shift + nonsense
645	exon 14	1685del4	PTC + 7 aa	shift + nonsense
029	intron 4	808-2 a → g	?	splice site mutation
162	intron 6	1129+2 t → g	?	splice site mutation
125	intron 7	1223+1 g → t	?	splice site mutation
143	intron 8	1299+1 g → a	?	splice site mutation
1620	intron 11	1538+5 g → a	(PTC + 6 aa)	splice site mutation
1006	intron 11	1538+3 del4	?	loss of exon 11 + shift
1605	intron 13	1661+1 g → t	?	splice site mutation
1012	intron 13	1662-2 a → t	?	splice site mutation
1626	intron 15	1812+1 g → a	?	splice site mutation
2992	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	splice site mutation
5226	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5330	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
1611	intron 16	1813-2 a → g	?	loss of exon 16 + shift
		1853+1 g → a		splice site mutation

^a The nt positions refer to the sequence of the SPG4 cDNA. ^b The aa positions refer to the spastin sequence. The exon bases are indicated in upper case, those of the introns in lower case. PTC+n aa - "premature termination codon" at n amino acids downstream of the mutation.

Example 5 : Analysis of the protein sequence of spastin

The open reading frame of SPG4 encodes a 616 aa protein which we have named spastin and the molecular weight of which is approximately 67.2 kDaltons (kD). The comparison of this amino acid sequence in the protein databases, using the BLAST programs, made it possible to reveal a region of strong homology with several members of the AAA family, at the C-terminal end of spastin. The "typical" motifs of the AAA family, encompassed in the AAA cassette, are located between aa positions 342 and 599 (see fig. 2) according to the sequence comparisons in the ProDom and Prosite protein domain databases. The three conserved typical domains, including the Walker A and B motifs and also the minimum consensus motif of the AAA proteins are located in the AAA cassette at aa positions 382-389, 437-442 and 480-498, respectively, (fig. 2). The Walker A motif, "GPPGNGKT", also called p-loop, which corresponds to the ATP-binding domain, and the B motif, "IIFIDE", are very conserved among all the members of the AAA family, including spastin.

The comparison of the AAA cassettes present in 150 proteins of this ATPase family, derived from organisms which are very far apart in evolution made it possible to classify this set of proteins into several subgroups, as a function of the number of AAA cassettes identified (1 or 2) and of the sequence homologies between these various cassettes (Beyer, 1997). Among all the proteins of the AAA family, spastin shows stronger homology with a particular subclass of the AAAs, and more specifically with the following proteins, most of which were identified through the complete sequencing of the genome of the organism in question: two proteins of *Caenorhabditis elegans*, O16299 and Q18128; two subunits of the 26S proteasome of *Saccharomyces cerevisiae*, Yta6p (Q02845) and TBP6 (P40328) (Schnall et al., 1994); a subunit of the proteasome of *Schizosaccharomyces pombe* (O43078); the SAP1 (P39955) and END13 (P52917) proteins of *S. cerevisiae* and the murine SKD1 protein (P46467) (Perier et al., 1994). The multiple alignment of these 8 proteins with spastin is represented in fig. 4A. Of the 257 amino acids encompassing the AAA cassette (aa positions 342-599), spastin shows 52%, 51% and 50% sequence identity with the Yta6p (Q02845) yeast protein, the O16299 nematode protein and the TBP6 (P40328) yeast protein, respectively. Similar results were obtained by analyzing the protein sequence of spastin in the ProDom database, which showed the existence of three domains of homology (named 92, 179 and 6226, and corresponding to aa positions 342-409, 411-509 and 512-599) found in the putative subunits of the 26S proteasome of yeast. In addition, the members of this AAA subgroup most commonly contain motifs of the leucine-zipper type, two of which could be detected

in the protein sequence of spastin at aa positions 50-78 and 508-529, by analyzing the sequence in the Prosite database (see fig. 2). This analysis was also able to predict the presence of a dimerization motif of the helix-loop-helix type, located between aa positions 478 and 486.

5 The comparison of the protein sequence of spastin with those of the mitochondrial metalloproteases, such as the AFG3, RCA1 and YME1 yeast proteins, and also paraplegin, which is implicated in a rare form of AR-HSP, shows that the homology between these five members of the AAA family is limited to the 257aa region encompassing the AAA cassette (fig. 4B). In this region, the sequence identity between
10 spastin and paraplegin is only 29%, whereas paraplegin and the AFG3 yeast protein are 57% identical over this same portion of the protein sequence. This sequence comparison suggests that spastin does not belong to the same AAA subgroup as paraplegin and other mitochondrial metalloproteases. In addition, the computer analysis of the spastin sequence using the PSORT II program, which makes it possible to predict the subcellular
15 location of the proteins, appears to indicate that spastin is a nuclear protein. A possible nuclear localization signal (NLS), RGKKK, was revealed between aa positions 7 and 11, whereas no signal peptide characteristic of importation into mitochondria could be detected, unlike what had been observed for paraplegin.

Example 6 : Expression profiles for SPG4 and for its murine ortholog Spg4

20 The comparison of the nucleic acid sequence of SPG4 in the EST databanks made it possible to detect several human, murine and rat ESTs showing strong homology with SPG4. The mouse blastocyst and E8 embryo cDNA clones corresponding to two of the murine ESTs, AA560327 and AA107866, were obtained from the IMAGE consortium and entirely sequenced. The assembly of the sequences of these cDNA clones made it
25 possible to reconstitute a 1689 bp consensus sequence including a 1514 bp incomplete open reading frame. The comparison between the human SPG4 cDNA and this mouse cDNA showed that the murine transcript lacks approximately 460 bp at the 5' end, including the translation initiation codon. The mouse open reading frame is followed by a 175 bp 3' noncoding region (3' UTR) containing a polyadenylation site located ~20 bp
30 upstream of the polyA tail (fig. 5). The nucleic acid sequence of SPG4 and the protein sequence of human spastin show 89% (between nt positions 460 and 1982) and 96% (between aa positions 113 and 616) identity, respectively, with the mouse cDNA and deduced protein sequences. This considerable degree of homology makes it possible to affirm that this mouse transcript corresponds to the murine ortholog of SPG4, which was
35 therefore named Spg4.

The hybridization of Northern blots comprising the mRNAs of various human and murine tissues (Clontech) with the SPG4 and Spg4 cDNA clones did not give any convincing results, except a very weak band corresponding to a 2.5 kb transcript in the mouse testicle after exposure for 10 days. Because of the low level of expression of this gene, the expression profiles for SPG4 and Spg4 were determined by PCR experiments on normalized collections of cDNA originating from various adult and fetal tissues (see fig. 6A to 6C). The murine Spg4 gene is expressed ubiquitously in the adult tissues of mice, and also from the E7 stage to the E17 stage of mouse embryos (fig. 6A). Higher expression of Spg4 was detected in the liver, skeletal muscle and testicles, and also at the E15 stage of embryos. The early expression of Spg4 during embryonic development was confirmed by the presence of ESTs originating from blastocyst, E8 embryo and embryonic carcinoma cDNA libraries in the public EST databanks. The human SPG4 gene is, itself, also expressed ubiquitously in adult (fig. 6B) and fetal (fig. 6C) tissues, with perhaps more marked expression in fetal brain.

Example 7 : No oxidative phosphorylation impairment in SPG4 locus-linked AD-HSP

In order to determine whether spastin mutations induced an oxidative phosphorylation (OXPHOS) impairment in mitochondria, in the same way as had been observed for paraplegin, a muscle biopsy was performed on a patient from one of the SPG4 locus-linked AD-HSP families. The morphological and histo-enzymatic analyses of this muscle biopsy did not reveal any muscle fibers of the RRF (for "ragged red fiber") type, characteristic of OXPHOS impairments in mitochondria. The fact that all the muscle fibers appear to be normal, and also the prediction of a nuclear localization for spastin, seem to indicate that SPG4 locus-linked AD-HSP is not a mitochondrial disease of the OXPHOS type, unlike SPG7 locus-linked AR-HSP.

Using a positional cloning approach based on sequencing a 1.5 Mb region, we have identified the SPG4 (or SPAST) gene responsible for the most common form of AD-HSP, previously located on chromosomal bands 2p21-p22. Thirty nine mutations which modify or are likely to modify the gene product, named spastin, could be detected in the affected individuals from forty one families with AD-HSP showing a link to the SPG4 locus. Spastin is a novel member of the AAA protein family, which appears to have a nuclear localization and which shows strong homology with the subunits of the 26S proteasome of yeast. Despite great homology restricted to a domain of 230 to 250 aa, termed AAA cassette, the many members of this protein family can participate in very varied cellular mechanisms, such as the transport of proteins in vesicles, cell cycle

regulation, organelle biogenesis, i.e. control of transcription, etc. However, all these cellular mechanisms involve the assembly, the functioning or the degradation of protein complexes, which suggest that the members of the AAA family are so-called "chaperon" proteins.

References

- Barany, F., (1991), *Proc. Natl. Acad. Sci. USA*, 88, 189-193.
- Beyer, A. Sequence analysis of the AAA protein family. *Protein Sci.* 6, 2043-2058 (1997).
- Bodansky M., *Principles of peptide synthesis*, (1984).
- 5 Bruyn, R.P.M. & Scheltens, P.H. Hereditary spastic paraparesis (Strumpell-Lorrain) in *Handbook of clinical neurology* Vol. 15 (ed. de Jong, J.M.B.V.) 301-318 (Elsevier Science Publishers B.V., 1991).
- Buckholz, R.G. *Curr. Op. Biotechnology* 4 : 538-542, 1993.
- Burg, J.L. et al. (1996), *Mol. and Cell. Probes*, 10, 257-271.
- 10 Carter, B.J. *Curr. Op. Biotechnology* 3 : 533-539, 1993.
- Casari, G. et al. Spastic paraplegia and OXPHOS impairment caused by mutations in Paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* 93, 973-983 (1998).
- Cherif D., Julier, C., Delattre, O., Derré, J., Lathrop, G.M., and Berger, R. *Proc. Natl. Acad. Sci. USA*. 87 : 6639-6643, 1990.
- 15 Chu, B.C.F. et al. (1986), *Nucleic Acids Res.*, 14, 5591-5603.
- Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billault, A., Guasconi, G., Gervy, P., Le Chumakov, I.M., Rigault, P., Le Gall, I., et al. *Nature* 377 : 175-183, 1995.
- Confalonieri, F. & Duguet, M. A 200-amino acid ATPase module in search of a basic function. *BioEssays* 17, 639-650 (1995).
- 20 Duck, P. et al. (1990), *Biotechniques*, 9, 142-147.
- Durr, A. et al. Phenotype of autosomal spastic paraplegia linked to chromosome 2. *Brain* 119, 1487-1496 (1996).
- Edwards, C.P., and Aruffo, A. *Curr. Op. Biotechnology* 4 : 558-563, 1993.
- Epstein, A. *Médecine/Sciences* 8 : 902-911, 1992.
- 25 Erlich, H.A., (1989), New York : Stockton Press.
- Guatelli J.C. et al. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878, 1990 et al. *Cell* 85 : 281-290, 1996.
- Fink, J.K. et al. Autosomal dominant familial spastic paraplegia : tight linkage to chromosome 15q. *Am. J. Hum. Genet.* 56, 188-192 (1995).
- 30 Hazan, J., Lamy, C., Melki, J., Munnich, A., de Recondo, J., & Weissenbach, J. Autosomal dominant familial spastic paraplegia is genetically heterogeneous and one locus maps to chromosome 14q. *Nature Genet.* 5, 163-167 (1993).
- Hazan, J. et al. Linkage of a new locus for autosomal dominant familial spastic paraplegia to chromosome 2p. *Hum. Mol. Genet.* 3, 1569-1573 (1994).

- Hedera, P. et al. Novel locus for autosomal dominant hereditary spastic paraplegia, on chromosome 8q. *Am. J. Hum. Genet.* 64, 563-569 (1999).
- Heinzlef, O. et al. Mapping of a complicated familial spastic paraplegia to locus SPG4 on chromosome 2p. *J. Med. Genet.* 35, 89-93 (1998).
- 5 Hentati, A. et al. Linkage of a locus for autosomal dominant familial spastic paraplegia to chromosome 2p markers. *Hum. Mol. Genet.* 3, 1867-1871 (1994).
- The Hereditary Spastic Paraplegia Working Group. Hereditary spastic paraplegia : advances in genetic research. *Neurology* 46, 1507-1514 (1996).
- Innis, M.A. et al. (1990), Academic Press.
- 10 Jouet, M. et al. X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the L1 gene. *Nature Genet.* 7, 402-407 (1994).
- Kievitis, T. et al. (1991), *J. Virol. Methods*, 35, 273-286.
- Köhler et Milstein. *Nature* 256, 495-497, 1975.
- Kwoh, D.Y. et al. (1989), *Proc. Natl. Acad. Sci. USA*, 86, 1173-1177.
- 15 Landegren U., Kaiser R., Sanders J. & Hood L. *Science* 241 : 1077-1080, 1988.
- Lizardi, P.M. et al. (1988), *Bio/technology*, 6, 1197-1202.
- Luckow, V.A. (1993), *Curr. Op. Biotechnology* 4, 564-572.
- Matthews, J.A. et al. (1988), *Anal. Biochem.*, 169 : 1-25.
- Miele, E.A. et al. (1983), *J. Mol. Biol.*, 171 : 281-295.
- 20 Nielsen, J.E. et al. CAG repeat expansion in autosomal dominant pure spastic paraplegia linked to chromosome 2p21-24. *Hum. Mol. Genet.* 6, 1811-1816 (1997).
- Olins, P.O., and Lee, S.C. *Curr. Op. Biotechnology* 4 : 520-525, 1993.
- Osoegawa, K. et al. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52, 1-8 (1998).
- 25 Perier, F. et al. Identification of a novel mammalian member of the NSF/CDC48p/Pas1p/TBP-1 family through heterologous expression in yeast. *FEBS lett.* 351, 286-290 (1994).
- Perricaudet, M., Stratford-Perricaudet, L. and Briand, P. *La Recherche* 23 : 471-473, 1992.
- 30 Polo, J.M., Calleja, J., Combarros, O. & Berciano, J. Hereditary ataxias and paraplegias in Cantabria, Spain. An epidemiological and clinical study. *Brain* 114, 855-866 (1991).
- Reid, E. Pure hereditary spastic paraplegia. *J. Med. Genet.* 34, 499-503 (1997).
- Rohlmann, A., Gotthardt, M., Willnow, T.E., Hammer, R.E., and Herz, J. *Nature Biotech.* 14 : 1562-1565, 1996.
- 35 Rolfs, A. et al. (1991), Berlin : Springer-Verlag.

- Sambrook, J., Fritsch, E.F., and Maniatis, T. Molecular cloning : a laboratory manual.
- Saugier-Weber, P. et al. X-linked spastic paraplegia and Pelizaeus-Merzbacher disease are allelic disorders at the proteolipid protein locus. *Nature Genet.* 6, 257-262 (1994).
- Schnall, R. et al. Identification of a set of yeast genes coding for a novel family of putative
- 5 ATPases with high similarity to constituents of the 26S protease complex. *Yeast* 10, 1141-1155 (1994).
- Scott, W.K. et al. Locus heterogeneity, anticipation, and reduction of the chromosome 2p minimal candidate region in autosomal dominant familial spastic paraplegia. *Neurogenetics* 1, 95-102 (1997).
- 10 Sec. Ed. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
- Segev, D., (1992), Kessler C. Springer Verlag, Berlin, New-York, 197-205.
- Skre, H. Hereditary spastic paraplegia in Western Norway. *Clin. Genet.* 6, 165-183 (1974).
- Stone, B.B. et al. (1996). *Mol. and Cell. Probes*, 10 : 359-370.
- Stewart J.M. et Yound J.D., solid phase peptides synthesis, Pierce Chem. Company,
- 15 Rockford, 111, 2ème éd., (1984).
- Suggs S.V., Wallace R.B., Hirose T., Kawashima E.H. and Itakura K. *PNAS* 78 : 6613-6617, 1981.
- Temin, H.M. Retrovirus vectors for gene transfer. In Kucherlapati R., ed. *Gene Transfer*, New York, Plenum Press, 149-187, 1986.
- 20 Walker G.T., Fraiser M.S., Schram J.L., Little M.C., Nadeau J.G., & Malinowski D.P. *Nucleic Acids Res.* 20 : 1691-1696, 1992.
- Werderlin, L. Hereditary ataxias. Occurrence and clinical features. *Acta Neurol. Scand.* 73 (Suppl. 106) (1986).
- Woo S.L.C. *Methods Enzymol.* 68 : 389, 1979.

CLAIMS

1. Purified or isolated nucleic acid of the SPG4 gene, characterized in that it comprises a sequence chosen from the group comprising:

- 5 a) the sequence SEQ ID No. 1, the sequence SEQ ID No. 2, the sequence SEQ ID No. 72, the sequence SEQ ID No. 106 or the sequence of at least 15 consecutive nucleotides of one of these sequences;
- b) the nucleic acid sequences which are homologs or variants of the sequences SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 or SEQ ID No. 106; and
- 10 c) the complementary sequence or the RNA sequence corresponding to the sequences as defined in a) and b).

2. Purified or isolated nucleic acid according to claim 1, with the exception of the nucleic acid identified in the GenBank databank under the accession number AB029006.

- 15 3. Purified or isolated nucleic acid according to claim 1 or 2, characterized in that it comprises at least one sequence of at least 15 consecutive nucleotides of the nt 714-809, ends inclusive, fragment of the sequence SEQ ID No. 2, of the sequence complementary thereto or of the sequence of the corresponding RNA thereof.

4. Purified or isolated nucleic acid according to one of claims 1 to 3, characterized in that it comprises a mutation corresponding to a natural polymorphism in humans.

5. Probe or primer, characterized in that it comprises a sequence of a nucleic acid according to one of claims 1 to 4.

6. Probe or primer according to claim 5, characterized in that its sequence is chosen from the sequences SEQ ID No. 4 to SEQ ID No. 71.

7. Splice acceptor or donor site, characterized in that it comprises a sequence of a nucleic acid according to claim 1 chosen from the sequences SEQ ID No. 74 to SEQ ID No. 105.

8. Method for screening cDNA or genomic DNA libraries, or for cloning isolated genomic or cDNA encoding spastin, characterized in that it uses a nucleic acid sequence according to one of claims 1 to 7.

9. Method according to claim 8, for identifying the genomic or cDNA sequence of the SPG4 gene of mammals, in particular of mice.

10. Method for identifying a mutation carried by the human SPG4 gene, characterized in that it uses a nucleic acid sequence according to one of claims 1 to 7.

11. Method according to claim 10, for identifying a mutation responsible for autosomal dominant hereditary spastic paraplegia.

12. Method for identifying the nucleic acid sequences which promote and/or regulate the expression of the SPG4 gene, characterized in that it uses a nucleic acid sequence according to one of claims 1 to 7.

13. Nucleic acid identified using a method according to one of claims 9 to 12.

14. Polypeptide encoded by a nucleic acid according to one of claims 1 to 4 and 13.

15. Polypeptide according to claim 14, preferably with the exception of the 584 amino acid peptide, the sequence of which is identified in the GenBank databank under the accession number AB029006.

16. Polypeptide according to claim 14 or 15, characterized in that it comprises an amino acid sequence chosen from the group comprising:

a) the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107 or the sequence of at least 10 consecutive amino acids of one of these sequences; and

b) the sequences which are homologs or variants of the sequences SEQ ID No. 3, SEQ ID No. 73 or SEQ ID No. 107.

17. Polypeptide according to claim 14 or 15, characterized in that it comprises the sequence of at least 8 consecutive amino acids of the sequence of the aa 197-228, ends inclusive, fragment of the sequence SEQ ID No. 3.

18. Polypeptide according to claim 14 or 15, characterized in that it comprises an amino acid sequence chosen from the group comprising the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107, which sequences carrying at least one of the mutations corresponding to a natural polymorphism in humans, and the sequences of the fragments thereof of at least 10 consecutive amino acids.

19. Cloning and/or expression vector containing a nucleic acid sequence according to one of claims 1 to 4, and 13.

20. Vector according to claim 19, characterized in that it includes the elements required for its expression in a host cell.

21. Host cell transformed with a vector according to claim 19 or 20.

22. Mammal, except a human, characterized in that it comprises a cell according to claim 21.

23. Mammal, except a human, according to claim 22, comprising a transformed cell, characterized in that the sequence of at least one of the two alleles of the SPG4 gene

contains at least one of the mutations corresponding to a natural polymorphism in humans or identified using a method according to claim 10 or 11.

24. Use of a nucleic acid sequence according to one of claims 5, 6 and 13, as a probe or primer, for detecting and/or amplifying nucleic acid sequences.

5 25. Use of a nucleic acid sequence according to one of claims 1 to 7, and 13, for screening a genomic or cDNA library.

26. Use of a nucleic acid sequence according to one of claims 1 to 4 and 13, for producing a recombinant or synthetic polypeptide.

10 27. Method for producing a recombinant polypeptide, characterized in that a transformed cell according to claim 21 is cultured under conditions which allow the expression of said recombinant polypeptide, and in that said recombinant polypeptide is recovered.

28. Polypeptide, characterized in that it is obtained using a method according to claim 27.

15 29. Mono- or polyclonal antibodies or their fragments, chimeric antibodies or immunoconjugates, characterized in that they are capable of specifically recognizing a polypeptide according to one of claims 14 to 18, and 28.

30. Method for detecting and/or purifying a polypeptide according to one of claims 14 to 18, and 28, characterized in that it uses an antibody according to claim 29.

20 31. Method for genotypic diagnosis of AD-HSP associated with the SPG4 gene, characterized in that a nucleic acid sequence according to one of claims 1 to 7 and 13 is used.

25 32. Method for genotypic diagnosis of AD-HSP associated with the presence of at least one mutation on a sequence of the SPG4 gene, using a biological sample from a patient, characterized in that it includes the following steps:

- a) where appropriate, isolation of the genomic DNA from the biological sample to be analyzed, or production of cDNA from the RNA of the biological sample;
- b) specific amplification of said DNA sequence of the SPG4 gene likely to contain a mutation, using primers according to either of claims 5 and 6 or a nucleic acid
- 30 according to claim 13;
- c) analysis of the amplification products obtained and comparison of their sequence with the corresponding normal sequence of the SPG4 gene.

33. Method for diagnosing AD-HSP associated with abnormal expression of a polypeptide encoded by the SPG4 gene, characterized in that one or more antibodies

35 according to claim 29 is (are) brought into contact with the biological material to be tested,

under conditions which allow the possible formation of specific immunological complexes between said polypeptide and said antibody or antibodies, and in that the immunological complexes possibly formed are detected and/or quantified.

34. Method for selecting a chemical or biochemical compound which is capable
5 of interacting directly or indirectly with a polypeptide according to one of claims 14 to 18, and 28, or with a nucleic acid according to one of claims 1 to 7, and 13, and/or which makes it possible to modulate the expression or the activity of these polypeptides, characterized in that it comprises bringing a nucleic acid sequence according to one of
10 claims 1 to 7, and 13, a polypeptide according to one of claims 14 to 18, and 28, a vector according to either of claims 19 and 20, a cell according to claim 21, a mammal according to either of claims 22 and 23 or an antibody according to claim 29 into contact with a candidate compound, and detecting a modification of the activity of said polypeptide.

35. Use of a nucleic acid sequence according to one of claims 1 to 7, and 13,
15 of a polypeptide according to one of claims 14 to 18, and 28, of a vector according to either of claims 19 and 20, of a cell according to claim 21, of a mammal according to either of claims 22 and 23 or of an antibody according to claim 29, for studying the expression or the activity of the SPG4 gene.

36. Kit or pack for diagnosis, characterized in that it comprises at least one
20 compound chosen from the following group of compounds:
a) a nucleic acid according to either of claims 5 and 6; and
b) an antibody according to claim 29.

**CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE
RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL DOMINANT
SPASTIC PARAPLEGIA**

ABSTRACT OF THE DISCLOSURE

5 The invention concerns the identification and characterization of the SPG4 gene encoding spastin, and some mutations thereof responsible for the most frequent form of autosomal dominant familial spastic paraplegia, to the cloning and characterization of its cDNA and the corresponding polypeptides. The invention also concerns vectors, transformed cells and transgenic animals as well as diagnostic methods and kits, and methods for selecting a chemical or biological compound capable of directly or indirectly interacting with said polypeptide.

1/7

FIGURE 1A

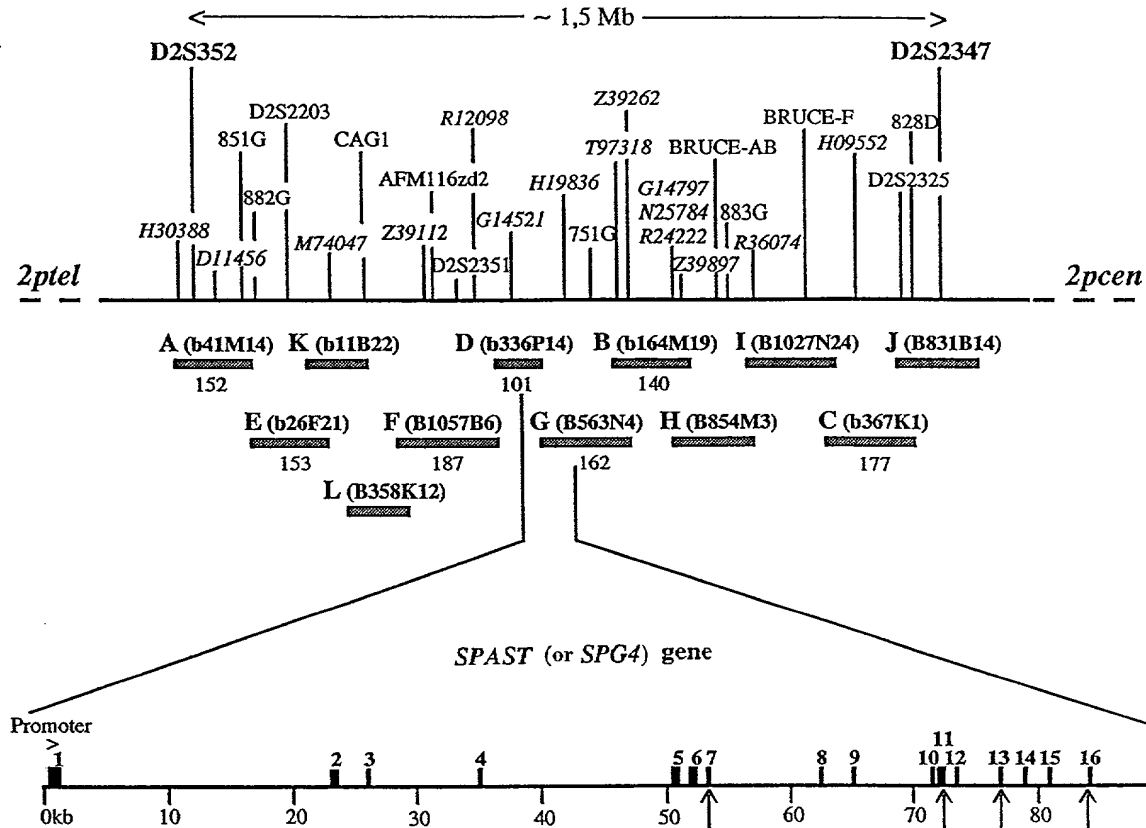
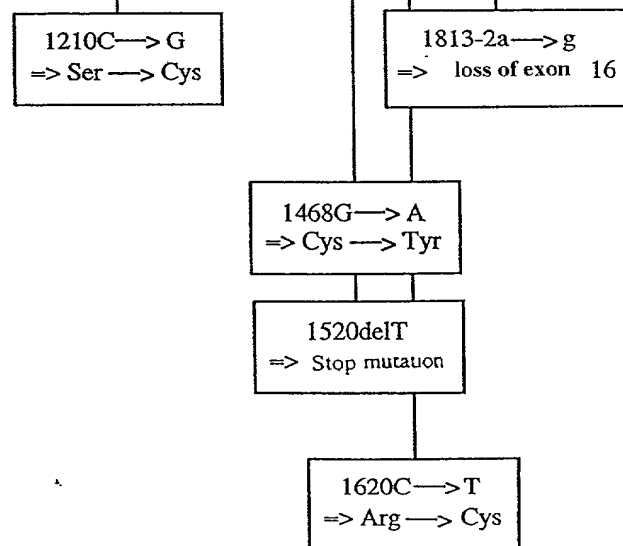


FIGURE 1B

FIGURE 1C



09/830902

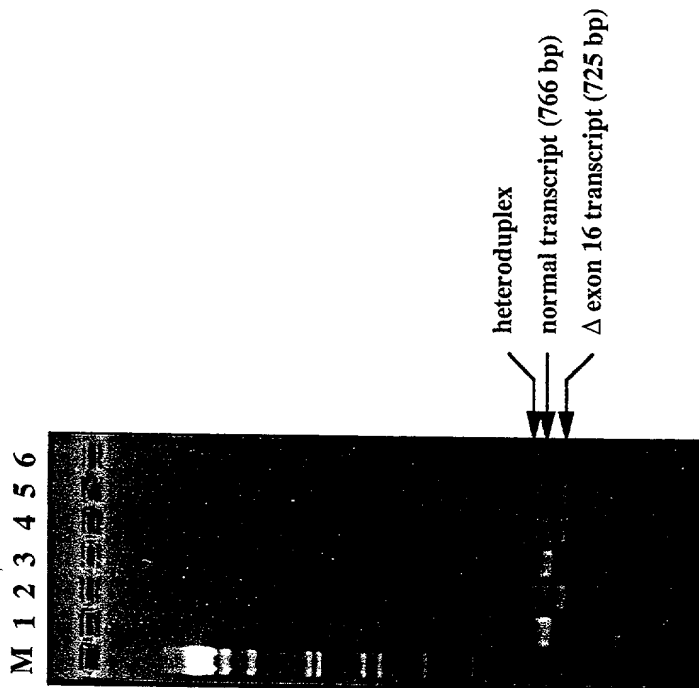
2/7

1 GCTCCTGAGACCGCGCGGCACACGGGGGTCTGTGGCCCCCGCGTAGCAGTGGCTGCGCGCCGTCGCTTGGTTCCCGTCGGTCTGCGGGAGGCGGG 95
1 TTATGGCGCGCGCGCGCAGTGAGAGCTGTGAATGAAATTCCTCCGGGTGGACGAGGGAAGAAAGGCTCCGGCGCGCCAGCAACCCGGTGCCTCC 190
1 M N S P G G R G K K K G S G G A S N P V P P
CAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCTCCCGCGCGCGGGCGCGCCCTCCGCGCGAGTCGCGGCATAAGCGGAACCTGTACTATTTCT 285
23 R P P P P C L A P A P P A A G P A P P P E S P H K R N L Y Y F S
CCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTTCGCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCC 380
55 Y P L F V G F A L L R L V A F H L G L L F V W L C Q R F S R A
CTCATGGCAGCCAAGAGGAGCTCCGGGGCGCGCCAGCACCTGCCTCGGCTCGGCCCCGCGCGCGTGGCGGGCGCGAGGCCGAGCGCTCCG 475
86 L M A A K R S S G A A P A P A S A S A P A P V P G G E A E R V R
AGTCTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGTGGACAGAAAGGAGCTAGACGCTTCAAGCTAAA 570
118 V F H K Q A F E Y I S I A L R I D E D E K A 2 G Q K E Q A V E W Y
ATAAGAAAGGTATTGAAGAACTGGAAAAAGGAATAGCTGTATAGTTACAGGACAAGTGAACAGTGTGAAGAGCTAGACGCTTCAAGCTAAA 665
150 K K G I E E L E K G I A V I V T G Q G 3 E Q C E R A R R L Q A K
ATGATGACTAATTTGGTTATGGCCAAGGACCGCTTACAACTTCTAGGAAGATGCAACAGTTTGGCATTTCCTCAAGTACACAAACGGACGCTA 760
181 M M T N L M A K D R L Q L L E K M Q P V L P F S K S Q T D V Y
TAATGACAGTACTAATTTGGCATGCCCAATGGACATCTCCAGTCACTAAGTGGAGCTGTTCCAAAAAGAAAAGACCCCTTAACACACACTAGTA 855
213 N D S T N L A C R N G H L Q S E S G A V P K R K D P L T H T S N
ATTCAGTCCCTCGTTCAAAACAGTATGAAACTGGATCTGCAGGCCTTTCAGGCCACCATAGAGCACCTAGTTACAGTGGTTTATCCATGGTT 950
245 S L P R S K T V M K T G S A G L S G H H R A P S Y S G L S M V
TCTGGAGTGAACAGGGATCTGGTCTGCTCCACTACATAAGTCTACTCCGAAAAACAAATAGGACAAATAAACCTTCTACCCCTACAACCTGC 1045
276 S G V K Q G S G P A P T T H K G 6 T P K T N R T N K P S T P T A
TACTCGTAAGAAAAAGACTTGAAGAATTTAGGAATGTGGACAGCAACCTTGTCTAACCTTATAATGAATGAAATTTGGACATTTGGACAGCTG 1140
308 T R K K K D L K N F R N V D S N L A N L I M N E I V D N G 7 T A V
TTAAATTTGATGATATAGCTGGTCAAGACTTGGCAAAACAGCATTTGCAAGAAATTTGTTATTCTTCTCTGAGGCGCTGATTTGTTTACAGGG 1235
340 K F D D I A G Q D L A K Q A L Q E I V I L P S L R P L F T G
CTTAGAGCTCTGCCAGAGGCTGTACTCTTTGGTCCACCTGGGAATGGGAAGACAATGCTGCTAAAGCAGTAGCTGCAGAATCGAATGCAAC 1330
371 L R A P A R G L L L F G P P G N G K T M L A 9 K A V A A E S N A T
CTTCTTTAATAAAGTGTGCAAGTTTAACTTCAAAATATCTGGGAGAGGAGAGAAATTTGGTGGAGGCTCTTTTGTGCTGGCTCGAGAACTT 1425
403 F F N I S A S L T S K Y V 10 G E G E K L V R A L F A V A R E L Q
AACCTTCTATAATTTTATAGTGAAGTTGATAGCCTTTTGTGTGAAAGAAGAGAAGGGGAGCAGATGCTAGTAGACGCTTAAACCTGAATTT 1520
435 P S I I F I D I I E V D S L L C E R R E G E H D A S R R L K T E F
CTAATAGAATTTGATGGTTTACAGTCTGCTGGAGATGACAGAGTACTTGTAAATGGGTGCAACTAATAGGCCACAAGAGCTTGTAGAGGCTGTTCT 1615
466 L I E F D G V 12 Q S A G D R V L V M G A T N R P Q E L D E A V L
CAGCGTTTCATCAAAACGGGTATATGTGCTTTTACCAATAGGAGACAAGACTACTTTTGTCTTAAATCTGTTATGTAACCAAGGAAGTCCAT 1710
498 R 13 R F I K R V Y V S L P N E E T 14 R L L L L K N L L C K Q G S P L
TGACCCAAAAAGAACTAGCACAACTTGCTAGTATGACTGATGATACTCAGGAAGTGACCTAACAGCTTTGGCAAAAGATGCAGCACTGGGTCCT 1805
530 T Q K E L A Q L A R M 15 T D G Y S G S D L T A L A K D A A L G P
ATCCGAGTAACTAAACAGCAACAGGTGAAGAATATGTCTGCCAGTGAGTGAAGAAATATTGATTATCTGACTTCACTGAATCCTTGAAAAAAAT 1900
561 I R E 16 L K P E Q V K N M S A S E M 17 R N I R L S D F T E S L K K I
AAAAACGAGCGTCAGCCCTCAAACTTTAGAAGCGTACATCGTTGGAACAAGGACTTTGGAGATACCAGTGTAAAGAAATACCTTTGTAAACC 1995
593 K R S V S P Q T L E A Y I R W N K D F G D T T V *
TGCAGAACATTTTACTTAAAGAGGAAACACAAGATCTTCAATGAACGTCATCGGTACAGAAACAGCCTAAGTTTACAGGACTTTTGTAGTCT 2090
TACATATTTGTGCACCAAACTTGAAGATGAACGAGAAACAGACTTAAACAAATATACAATGCAATGTAATTTTGTGTTTAAAGGCCTTGC 2185
CTTGATGGTCACAGTTATCCCAATGGACACTAAGTTAGAGCACAAACAAACCTGATTCTGGTCTTCTTTACCAATATAATCATAATGTAAATAAT 2280
AATTTGTATATTGTGTGAGATGAAAGTATCCAGGAACAGTGAATGGTAGAAGACACAAGAACATTGTTGTTTGTCTTCTGATGTTTTTTC 2375
TTAAATAGTAATTTCTCTACTTTTCTTTTCTACTGTTGTCTTAACTACAGGTGATTGGAATGCCAAACACTCTTAAGTTTATTTCTTTTTC 2470
GTTTATAAATTCAGTGTGCCAAATGAACTTTTCTCTAAGTAACTGTAATAGGAAAAAGTTTATTTTGTAGAGTTTCTTCTTCAATAATCTACA 2565
GACATTAACAATTTGTGTGTTCTTTTACCTTTTATTTTCTATTACCTTGTACCAACAGTTTAGATAGCAATATAATAGCAAAAAAGCAAA 2660
TATGGTAAATAGAGAAGGTTTGAAGGTTTGAAGTTTACTCTGTCTATATAACATGTAGATCAGTCTTCAATGTGACCTGCAGTATTTTTTTTCTAAT 2755
GTATTTGTGAGAAATCTGTGTGAGACTGTTAACTTCTTCTGATGGAATTTATTTCTGCAAGAATTATCTGATATTAAAGAGAGCCAAATTTTA 2850
ACTGCTGTGAAATGTTTCCAGTGAAGAGAAGGAAAACTAGGAACATAAGACATTTCTAATTTATGCTTATTACTTTCTAATTTTACAGGA 2945
TAATTAAGCAAGTGGAACTACCATCTTTTATCTTAAATAATTATTAATCCCTTCAATGAACTTTAAAAAACTGAATTTTATACATGCGCAT 3040
ACATTTTCTAGTTCCTTCTGCTTGTCTTTATTAACCTCAAAAGTCTAGTCTAGTCTGTTGATCTGCCTTTTGTCTCCCAAAATGTACAGTAAT 3135
TCCATTTGTTGTATAAATATGCTGGATTTTCAATTAATAAATGTCTATTGTAGGGAGTAGAGACTCATATCATGGCCTTTTAAATATTGTAATA 3230
AAGGCAATAGATATTGCCCCCTAGTTTACTGG 3263

FIGURE 2

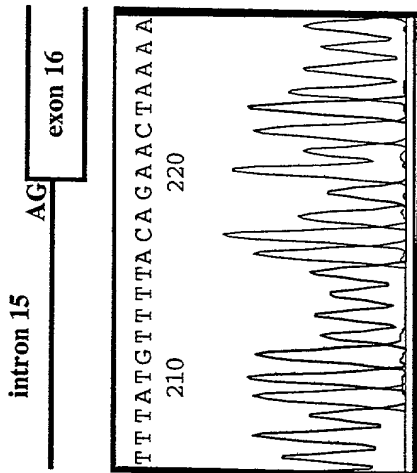
TTTATGTTTTCAGAACTAAA

A



B

Normal acceptor splice site



Mutated acceptor splice site

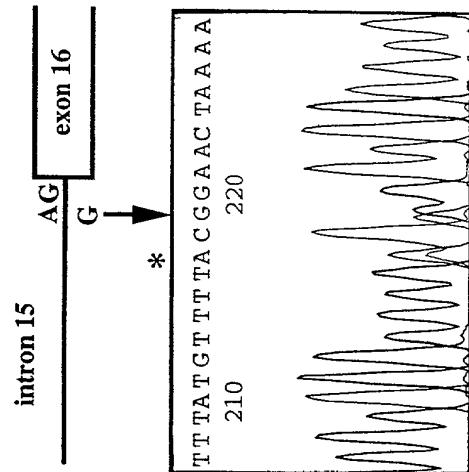


FIGURE 3

4/7

TBP6_YEAST -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI
Q02845_YEAST -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI
SAP1_YEAST MDSQSRHHILTRITKRRRPPQPLDTTELYSRIANETIYLLNLEKKRYKEALQGWKALTDTVLFFKQTLIEBNYPNTQSYTKDESVLQNGIREL
SKD1_MOUSE -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI
EN13_YEAST -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI
SPAST_HUMAN -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI
SPAST_MOUSE -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI
Q43078_SCHPO -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI
Q16299_CAEL -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI
Q18128_CAEL -----MAHEKFS-----IPEN-----FTLAQS-----LQLLYSVVKNQYKKNLADLIINSKGNKDTVTYCK-----IHKNLDTLLVYVN-----EGLRKI

TBP6_YEAST -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK
Q02845_YEAST -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK
SAP1_YEAST YHKSVMHLKRVKLVREEPAPRNDMPSSKTYTNHSSSTFTRSTFFVFQMVPGRMKTLRNANACCYKTAISNPSLSSYGNSTSIKRGEAENIR
SKD1_MOUSE -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK
EN13_YEAST -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK
SPAST_HUMAN -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK
SPAST_MOUSE -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK
Q43078_SCHPO -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK
Q16299_CAEL -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK
Q18128_CAEL -----EKTITLTKGCLGNLVVDHPELSIIE-----FOIL-----QODIRIARRKAETLMAEONGSPSLSSSSSVLGLGTGNGLRFPK

TBP6_YEAST -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE
Q02845_YEAST -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE
SAP1_YEAST VNFVPSKPLSNASRQHNKFNIEHNDPFLKKELEYSDKYISEFILIDLTNDEDDHDVGLKGNHVFDEESDGFEDVDYDNFSEVDVEE
SKD1_MOUSE -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE
EN13_YEAST -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE
SPAST_HUMAN -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE
SPAST_MOUSE -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE
Q43078_SCHPO -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE
Q16299_CAEL -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE
Q18128_CAEL -----LWRMGSKRDLKEADEKAKINKQADNIRARKLEEKKLGAKLQYERDLE-----LQREKLEIKVKEKVEFEVAQKLE-----EERVKEEE

TBP6_YEAST -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG
Q02845_YEAST -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG
SAP1_YEAST EKEERRRIKT-----LEAIQQQMSDLSTVSTSSSNKSVSSSENVPGSCIQSLPTAFALSLFPFPLNVDRASTGALKPHSLTSTMDSSKIRNP
SKD1_MOUSE -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG
EN13_YEAST -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG
SPAST_HUMAN -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG
SPAST_MOUSE -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG
Q43078_SCHPO -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG
Q16299_CAEL -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG
Q18128_CAEL -----ERKHR-----EQAEKKRISTLKHERKTN-----YKSRASLDNFSSSSSKSSGKIDN-----SLIKRRSLDVVSTSGEVRTP-----VRKSMMAAIG

TBP6_YEAST -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI
Q02845_YEAST -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI
SAP1_YEAST QISKLKHNHVVYLLKCTKSTPTLITKSTPTITKSKNTKPIIKSNAGSPSTSLVSNVSIQPKTAAMAAKRVLSNKKVWASPNALNTTKKSHPI
SKD1_MOUSE -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI
EN13_YEAST -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI
SPAST_HUMAN -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI
SPAST_MOUSE -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI
Q43078_SCHPO -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI
Q16299_CAEL -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI
Q18128_CAEL -----MAAQLAWS-----QYONGAN-----HSK-----VSNNGSHSNELQIR-----YKP-----TPPLKKRYDIKK-----PTVN-----RPI

TBP6_YEAST -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE
Q02845_YEAST -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE
SAP1_YEAST LKSKTAKVPNSSSK-----TSSHPSPRVSNSSKPYSGAAGSNKKPKSNOTTSMSKTNR-----KIPAKKIGSPKIEDVGTEDATEHATSLNEQREFEED
SKD1_MOUSE -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE
EN13_YEAST -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE
SPAST_HUMAN -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE
SPAST_MOUSE -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE
Q43078_SCHPO -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE
Q16299_CAEL -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE
Q18128_CAEL -----IKSPTLNRRQNSKSSNNIPTNSKLSKSKSNTNKVSRNEQNLEPS-----SPVLVSAT-----AVPAESK-----PMRSKSGTPDKESSASSSLDSRKE

TBP6_YEAST -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
Q02845_YEAST -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
SAP1_YEAST KKVLEILELEDEILDSLGVDRAAKQIFAEIIVVHGDEVDWD-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
SKD1_MOUSE -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
EN13_YEAST -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
SPAST_HUMAN -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
SPAST_MOUSE -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
Q43078_SCHPO -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
Q16299_CAEL -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA
Q18128_CAEL -----DILKSVQGVDRNACEQILNEILVTDEKVIWE-----LLEN-----NSK-----KAA-----VY-----FLR-----DL-----KL-----E-----VR-----M-----F-----T-----TMIAKA

TBP6_YEAST -----TESN-ST-----SV-----ASS-----L-----YL-----ES-----LVRA-----YM-----KKLSF-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
Q02845_YEAST -----TESN-ST-----SV-----ASS-----L-----YL-----ES-----LVRA-----YM-----KKLSF-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
SAP1_YEAST TESN-ST-----SV-----ASS-----L-----YL-----ES-----LVRA-----YM-----KKLSF-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
SKD1_MOUSE TEANNST-----SI-----SSD-----V-----WM-----ES-----LVKN-----OL-----RENKP-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
EN13_YEAST TEAN-ST-----SV-----SSD-----V-----WM-----ES-----LVKN-----OL-----RENKP-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
SPAST_HUMAN AASN-AT-----NI-----AAS-----T-----YV-----EG-----LVRA-----AV-----RELOP-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
SPAST_MOUSE AASN-AT-----NI-----AAS-----T-----YV-----EG-----LVRA-----AV-----RELOP-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
Q43078_SCHPO TEAK-AT-----SI-----ASS-----T-----YV-----EG-----LVRA-----AV-----RELOP-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
Q16299_CAEL GSCK-AT-----NI-----ASS-----T-----YV-----EG-----LVRA-----AV-----RELOP-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL
Q18128_CAEL GSCK-AT-----NI-----ASS-----T-----YV-----EG-----LVRA-----AV-----RELOP-----I-----I-----SMLTA-----SD-----NENESS-----I-----I-----L-----I-----QWSSSL

TBP6_YEAST -----SSATAQSE-----DRNNT-----LDSRV-----LG-----L-----WAI-----D-----AR-----SRKLYIP-----DYET-----LYHLKRLMAKQ-----KNSLQDLQYELITM-----E-----F-----GSS
Q02845_YEAST -----SSATAQSE-----DRNNT-----LDSRV-----LG-----L-----WAI-----D-----AR-----SRKLYIP-----DYET-----LYHLKRLMAKQ-----KNSLQDLQYELITM-----E-----F-----GSS
SAP1_YEAST SSATAQSE-----DRNNT-----LDSRV-----LG-----L-----WAI-----D-----AR-----SRKLYIP-----DYET-----LYHLKRLMAKQ-----KNSLQDLQYELITM-----E-----F-----GSS
SKD1_MOUSE GVDND-----I-----LG-----L-----WAI-----D-----AR-----SRKLYIP-----DYET-----LYHLKRLMAKQ-----KNSLQDLQYELITM-----E-----F-----GSS
EN13_YEAST GVDND-----I-----LG-----L-----WAI-----D-----AR-----SRKLYIP-----DYET-----LYHLKRLMAKQ-----KNSLQDLQYELITM-----E-----F-----GSS
SPAST_HUMAN QAGD-----RV-----MG-----E-----L-----E-----VL-----IKRVYVS-----HEET-----LLLLKNLLCKQ-----GSPLTQKELAQALARM-----D-----Y-----GSS
SPAST_MOUSE QAGD-----RV-----MG-----E-----L-----E-----VL-----IKRVYVS-----HEET-----LLLLKNLLCKQ-----GSPLTQKELAQALARM-----D-----Y-----GSS
Q43078_SCHPO THAPD-----HSPRV-----MG-----E-----L-----E-----VL-----IKRVYVS-----HEET-----LLLLKNLLCKQ-----GSPLTQKELAQALARM-----D-----Y-----GSS
Q16299_CAEL NTAPD-----HSPRV-----MG-----E-----L-----E-----VL-----IKRVYVS-----HEET-----LLLLKNLLCKQ-----GSPLTQKELAQALARM-----D-----Y-----GSS
Q18128_CAEL TSADD-----RI-----IG-----E-----L-----E-----VL-----IKRVYVS-----HEET-----LLLLKNLLCKQ-----GSPLTQKELAQALARM-----D-----Y-----GSS

TBP6_YEAST -----LTSIAKAE-----AME-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
Q02845_YEAST -----LTSIAKAE-----AME-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
SAP1_YEAST LTSIAKAE-----AME-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
SKD1_MOUSE LSIIVRD-----AMG-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
EN13_YEAST LSIIVRD-----AMG-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
SPAST_HUMAN LSIIVRD-----AMG-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
SPAST_MOUSE LSIIVRD-----AMG-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
Q43078_SCHPO LSIIVRD-----AMG-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
Q16299_CAEL LSIIVRD-----AMG-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS
Q18128_CAEL LSIIVRD-----AMG-----DLGDKLMFADF-----DKIRG-----IEIKRPFSECSVDNKKERVVRSLOKI-----I-----F-----GSS

TBP6_YEAST -----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
Q02845_YEAST -----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
SAP1_YEAST KWSQFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
SKD1_MOUSE KTFEDFGQEG-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
EN13_YEAST QTFEDFGQEG-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
SPAST_HUMAN RWNKDFGDTTV-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
SPAST_MOUSE RWNKDFGDTTV-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
Q43078_SCHPO EWNKDFGDTTV-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
Q16299_CAEL AWDKFGCLPPPSISR-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS
Q18128_CAEL DFRSFGC-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS-----EWSKFGSNGS

FIGURE 4A

5/7

```

AFG3_YEAST ---MMWQRYARGAPR-----SLTSL-----FGKASRISTVKP-----VLRSMRMPVRQRLQTLSGLATR-----NTIHR--S
RCA1_YEAST -MILLSWSRIATKVVRRPVRFRSYGLTHIKSLHIOYRLRLQENKSGKNEDNEDAKLNKEIPTDEVEAIRKQVEKYEIEQTKNNTIPANWK
PARAPLEGIN MAVLLLLLRLARRCPGCPGR-----PLNGPGPAWS-PGFPAPPCGCRP-----YMASRPPGD--LAEAGGRALQ-----S
YME1_YEAST ---MNVSKILVSPVTTN-----VLRIFAPRLPQIGASLLVQKKWA-----LRSKKFFRYFYSEKNSGMP-------K
SPAST_HUMAN --MNSPGGGRKKKGGSGGASN-----PVPPRPPPCCLAPAPPAAGPAPP-----ESPHKRNLYFYFYPLVFGFALLRLVAFHGLGLFVWLC

AFG3_YEAST TQIRSFHISWTRLN-----ENRPNKE---GEGKNNGKNKDNNSKEDGKDKRNEFGSL-----EYFRSKEFANTMTLTIGFTTIF
RCA1_YEAST EQKRKIDESIRLEDAVLKQESNRIOERKEKEENGPSKAKSNRTKEQGYFEGNNSRNIPPPPPPPKPLNDPSNPVSKNVNLFQIGLTFFL
PARAPLEGIN LQRLRLTPTTFEGINGLLKQH--LVQNPVRLWQLGGTFYFNTSRLKQKNEKDKSKGKAPFEDE-----EERRRRERDDQMYRERLRTLLV
YME1_YEAST KEADSSGKASNKST-----ISSIDNSQP--PPPSNTNDKTKQANVAVSHAMLATREQEA-----NKDLTSPDAQAAFYKLLQSNY
SPAST_HUMAN QRFSRALMAAKRSSGAAPAP--ASASAPAPVP--GGAERVRVFKQAFYISIALRIDEDEKAG-----QKEQAVEWYKKGIEELEKGIJAV

AFG3_YEAST TLTTPSSNNSGDDSNRVLTFQDFKTKYLEKGLVSKIY--NKF-LVEAELVN-----TKQVVSFTIGSV---DIFEEQ---MDQIQDILLNI
RCA1_YEAST LSFLLDLNLSLEEQS-EITWQDFREKLLAKGYVAKLIV--NKS-MVKVMLNDNGKNQADNYGRNFYFTIGSI---DSFEK--LQKAQDELDI
PARAPLEGIN IAVVMSLLNALSTSGGSISWNDFVHEMLAKGEVQVQV--PESDVVVEYLHPGAVVFGPRRLALMYRMQVANI---DKFEK--LRAADELNI
YME1_YEAST PQYVVSRTFETPGIASSPECMELYMEALQRIHSEADA--RQN-LTASSAG-----AVNPSLASSSNQ-----SGYHGN--FPSMYSPLYG
SPAST_HUMAN IYVQGQEQCERARRLQAKMMTNLVMADRLQLLEKMQP--LFFSKSQTDDVYNDSTN--LACRNGHLQSESGAVPRKRDPLHTSNTSLPSKSTVMKT

AFG3_YEAST PPRDRIPIKYIERSSPFTFLFFP---LPTIILLGGLYFITRKINSSPPNANGGGGGGGLGGMFNVGKSRAKLFNKETDIK---ISKNVA--CDE
RCA1_YEAST DKDFRIPVLYVQEGNWKAMFQI---LFTVLMIAIGIWLTRRS-----AQAGSGSGGIFGLSRKAKKKTETDVK---IK--KDVAC--CDE
PARAPLEGIN EAKDRIPVSYKRTGFFGNALYSV---GMTAVGLAILMYVFRLAG-----MTGREGGFSAFNQLKMARFTIYDGMKMGK---VS--KDVAC--CDE
YME1_YEAST SRKE--PLHVVSSESTFTVVSRW---VKWLLVFGILTYSFSEG-----FKYITENTTLKSEVADKSVDAKTN---VK--KDVAC--CDE
SPAST_HUMAN GSAGLSGHRAPSYSGLSMVSGVKQSGFAPPTHKGTETKTRNTKPTPTTATRKKKDLKNFRNVDNLANLIMNEVDNGTAVK--DDIA--QDL

AFG3_YEAST KQIMF--HFLKN--CKYTEKLGAKIPRAIS--T--L--K--T--C--ANVP--LSVSGSEFVMEFV--ASRV--D--TO--SMA--CIF--S
RCA1_YEAST KEIMF--SFLKE--SRYEKMGAKIPRAIS--T--L--K--T--C--ANVP--LSVSGSEFVMEFV--ASRV--D--TO--SMA--CIF--S
PARAPLEGIN KLEVR--FDYLS--ERFLQLGAKVPRALD--C--L--K--T--C--ANVP--LSVSGSEFVMEFV--ASRV--D--TO--SMA--CIF--S
YME1_YEAST RARLE--IDFLKD--TKYESLGKLPK--VLT--T--L--K--T--C--ANVP--LSVSGSEFVMEFV--ASRV--D--TO--SMA--CIF--S
SPAST_HUMAN KQALO--I--ILPSLR--ELFTGLRAPA--RLLF--N--M--K--K--V--A--SNAT--FNISAASLTSKYV--E--EKLVA--AV--ELO--SIF--S

AFG3_YEAST AIGKE--GKGALGGANDEREATLNQL--V--M--FTTS--DOVVVLAC--RPDI--K--L--P--G--FDRHIQIDS--DVNG--QOQYLVHLKRLNLDPLLT
RCA1_YEAST AIGKA--QKG--NFGSANDERENTLNQM--V--M--FTPA--DHVVVLAC--RPDI--K--L--P--G--FDRHIQIDS--DVNG--QOQYLVHLKRLNLDPLLT
PARAPLEGIN AVGKK--STT--MSGFSNTEEEQTLNQL--V--M--MGTT--DHVIVLAS--RADI--G--L--P--G--LDRHVFDL--TLQE--REIFEQLHLSKLKTQSST
YME1_YEAST AIGKK--NPK--DQAYAKQTLNQL--V--L--FSQT--SGIIIGA--FPEA--K--L--P--G--FDRHIQIDS--DVNG--QOQYLVHLKRLNLDPLLT
SPAST_HUMAN SLLCE--REG--EDASRRLKTET--I--F--VQSGAGDDRLVMGA--RPQE--E--V--L--F--FIKRVYVSL--NEET--LLLLKNLLCQGSPLTQK

AFG3_YEAST DDMNNLSGKL--TL--P--FT--ADIANACNE--LIAARHNDPYITIHFEO--IERVIAGLEKKTRV--SKEEKRSVAYHEAGH--VCGWFLKYADPLLV
RCA1_YEAST DLKN--RL--AL--P--FS--ADIANACNE--LIAARHNDPYITIHFEO--IERVIAGLEKKTRV--SKEEKRSVAYHEAGH--VCGWFLKYADPLLV
PARAPLEGIN FYSQ--RL--EL--P--FS--ADIANACNE--LIAAREGHTSVHTLNFEY--VERVLGATAKKSKI--SKEEQKVVAFHESGH--LVGMWLEHTEAVMKV
YME1_YEAST PTII--RG--PLS--AELANLVNO--VYACQKNVSVSDHFEW--KDKILMGAERKTMV--TDAARKATAFHEAGH--IMAKYTNATPLYKA
SPAST_HUMAN ELAQ--L--RM--D--YS--SDLTALAKD--LGPIRELKEPQVKN--MS--SEMRNIRLSDFTES--KKIKR--SVSPQPLE--YIRWNKDFGDTTV--

AFG3_YEAST SIIPRGQCALGYAQYLPDPQYLISEEQFRHRMIMALGGRVSEELHFP--SVTSGAHDDFKKVTQMANAMVTSLGMSPKIGYLSFDQNDG--NFKV
RCA1_YEAST SIIPRGQCALGYAQYLPDIFLLTEQQLKDRMTSLGGRVSEELHFP--SVTSGASDDFKKVTSMATAMVTELGMSPKIGWVNYQKRD--DSDL
PARAPLEGIN SITPRTNAAALGFAQMLPRDQHLFTKEQLFERMCALGGRASEALFN--EVTSGAQDDLKRVTRIAYSMVKQFGMAPGIGPISFPEAQEGLMGIG
YME1_YEAST TILFRGALGITFQLPEMDKVDITKRECQARLDVCMGGKIAEELIYKDNNTSGCGSDLSQATGTARAMVTQYGMSSDDVGPVNLSENW-----
SPAST_HUMAN -----

AFG3_YEAST NKPFSSNKTARTIDLEVSKIYVDDAARACTELLKKNLDKVDLVAKEKLLRKEATITREDMIRLLGFRP--FKERNEAFEKYLD-----PKSNTPE
RCA1_YEAST TKPFSDETCDIIDSEVYRIVQCHDRCTKLLKEKADVEKIAQVLLKKEVLTREDMIDLLCKRP--FPERNDAFDKYLNDYETEKIRKEEKNKR
PARAPLEGIN RRPFSQCLQOMMDHEARLLVAKAYRHTKVLQDNLDKQALANALLEKEVINVEDIEALIGPPPHGPKKMIAPQRWDIAQR--EKQDLGEEETET
YME1_YEAST -ESWSNKIRDIADNEVIELLKDSEERARRLLTKKNVELHRLAQGLIEYETLDAHEIEQVCKGEK--LDKLTSTNTVVEGPD--S--DERKIDGDOK
SPAST_HUMAN -----

AFG3_YEAST PEAPAAATN-----
RCA1_YEAST NEPKPSTN-----
PARAPLEGIN QQFPLGGEEETWEK-----
YME1_YEAST PKIPTMNA-----
SPAST_HUMAN -----

```

FIGURE 4B

Human: 1 [...] 459
Mouse: 1 AGGCCGAGAGCGTCCGCGTCTTCCACAAGCAGGCCCTTCGAGTACATCTCCATTGCCCCGTC 60
|||||
Human: 460 AGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCCTTCGAGTACATCTCCATTGCCCCGTC 519
Mouse: 61 GCATCGACGAGGAAGAGAAACAGGACAGAGGAACAGCTGTGGAATGGTATAAGAAAG 120
|||||
Human: 520 GCATCGATGAGGATGAGAAAGCAGGACAGAGGACAGCTGTGGAATGGTATAAGAAAG 579
Mouse: 121 GTATCGAAGAACTGGAAAAGGAATCGCTGTTATAGTTACGGGCCAAGGTGAACAGTATG 180
|||||
Human: 580 GTATTGAAGAACTGGAAAAGGAATAGCTGTTATAGTTACAGGACAAAGTGAACAGTGTG 639
Mouse: 181 AAAGAGCTAGACGCTCTCAAGCCAAAATGATGACTAATTAGTTATGGCCAAGGACCGTT 240
|||||
Human: 640 AAAGAGCTAGACGCTCTCAAGCTAAAATGATGACTAATTGGTTATGGCCAAGGACCGCT 699
Mouse: 241 TACAACCTCTAGAGAAGCTGCAACCAAGTTTGCATTTTCCAAAGTCACAGACGGACGCTCT 300
|||||
Human: 700 TACAACCTCTAGAGAAGATGCAACCAAGTTTGCATTTTCCAAAGTCACAAACGGACGCTCT 759
Mouse: 301 ATAACGAGAGTACTAAGCTGACATGCCGCAATGGACATCTCCAGTCAGAAAGTGGAGCAG 360
|||||
Human: 760 ATAATGACAGTACTAAGCTTGGCATGCCGCAATGGACATCTCCAGTCAGAAAGTGGAGCTG 819
Mouse: 361 TTCCGAAGAGGAAGACCCCTTAACACATGCTAGTAATTCATTGCCTCGATCAAAAACCTG 420
|||||
Human: 820 TTCCAAAAGAAAAGACCCCTTAACACACACTAGTAATTCAGTGCCTCGTTCAAAAACAG 879
Mouse: 421 TCCTGAAAAGTGGCTCCGACAGGCTCTCCGGTCACCAAGGGCCCTAGTTGCAGTGGTT 480
|||||
Human: 880 TTATGAAAAGTGGATCTGCAGGCCCTTCAGGCCACCATAGAGCACCTAGTTACAGTGGTT 939
Mouse: 481 TGTCCATGTTTCTGGAGCAGACCGGGACCTGGTCCCTGCAGCTACCAACATAAGGGTA 540
|||||
Human: 940 TATCCATGTTTCTGGAGTGAACAGGGATCTGGTCCCTGCTCCTACCACTATAAGGGTA 999
Mouse: 541 CTCCAAACCAAATAGAACCAACAAACCTTCTACTCCCAACACTGCAGTTCGGAAAAAGA 600
|||||
Human: 1000 CTCGAAAACAAATAGGACAAATAAACCTTCTACCCCTACAACTGCTACTCGTAAGAAAA 1059
Mouse: 601 AAGACTTGAAGAAATTTAGGAATGTGGACAGCAATCTTGCTAACCTTATAATGAATGAAA 660
|||||
Human: 1060 AAGACTTGAAGAAATTTAGGAATGTGGACAGCAACCTTGCTAACCTTATAATGAATGAAA 1119
Mouse: 661 TGTGTGACAAATGGGACAGCTGTTAAGTTTGATGACATAGCCGGCAGGAGCTGGCAAGC 720
|||||
Human: 1120 TTGTGGCAATGGAACAGCTGTTAAATTTGATGATATAGCTGGTCAAGACTTGGCAAAAC 1179
Mouse: 721 AAGCGCTGCAGGAGATTGTATCTCTCTCTCTGCGGCTGAGTTGTTACAGGGCTCA 780
|||||
Human: 1180 AAGCATTCAGAGAAATTTGTTATCTCTCTCTCTGAGGCTGAGTTGTTACAGGGCTTA 1239
Mouse: 781 GAGCTCCTGCTAGAGCTGTTACTCTTCGGTCCGCCAGGAAACGGAACCAATGCTGG 840
|||||
Human: 1240 GAGCTCCTGCCAGAGGGCTGTTACTCTTTGGTCCACCTGGGAATGGGAAGCAATGCTGG 1299
Mouse: 841 CTAAGCAGTAGCTGCAGAGTCTAATGCGACCTTTTCAACATAAGTCTGCCAGTTTAA 900
|||||
Human: 1300 CTAAGCAGTAGCTGCAGAAATCGAATGCAACCTTCTTTAATATAAGTCTGCAAGTTTAA 1359
Mouse: 901 CTTCAAAATATGTGGGAGAGGAGAGAAATTTGGTGAAGCTCTCTTTGCTGCTGCTCGAG 960
|||||
Human: 1360 CTTCAAAATACGTGGGAGAGGAGAGAAATTTGGTGAAGGCTCTTTTGTCTGCTCGAG 1419
Mouse: 961 AACTTCAACCATCTATAATTTTATAGATGAAGTTGACAGTCTTTTGTGTGAGAGACGGG 1020
|||||
Human: 1420 AACTTCAACCTTCTATAATTTTATAGATGAAGTTGATAGCCTTTTGTGTGAAGAAGAG 1479
Mouse: 1021 AAGGGGAGCAGCAGCTAGCAGACGGCTAAAGACGGAATTTTAAATAGAATTTGACGGGG 1080
|||||
Human: 1480 AAGGGGAGCAGATGCTAGTAGCGCTAAACCTGAATTTCTAATAGAATTTGATGGTG 1539
Mouse: 1081 TGCAATCTGCTGGAGATGACAGAGTACTTGAATGGGTGCAACTAACAGGCCCAAGAGC 1140
|||||
Human: 1540 TACAGTCTGCTGGAGATGACAGAGTACTTGAATGGGTGCAACTAATAGGCCACAAGAGC 1599
Mouse: 1141 TTGATGAAGCTGTTCTCAGCGCTTTCATTAACGGGTATATGTGCTTACCAAATGAGG 1200
|||||
Human: 1600 TTGATGAGGCTGTTCTCAGCGCTTTCATCAACGGGTATATGTGCTTACCAAATGAGG 1659
Mouse: 1201 AGACAAGACTCCTTCTGCTTAAACCTGTTGTGTAAACAAGGAAGTCCACTGACCCAA 1260
|||||
Human: 1660 AGACAAGACTCTTTTGTCTTAAACCTGTTTATGTAAACAAGGAAGTCCATTGACCCAA 1719
Mouse: 1261 AAGAAGCTCGCAGCTTGCTAGAAATGACCGATGGATACTCTGGAAGTGATCTGACCGCT 1320
|||||
Human: 1720 AAGAAGCTAGCACAACTTGCTAGAAATGACTGATGGATACTCAGGAAGTGACCTAACAGCTT 1779
Mouse: 1321 TGGCCAAGGATGACAGCCCTGGGTCTATCCGAGAAGCTGAAGCCAGAGCAGGTGAAGAATA 1380
|||||
Human: 1780 TGGCAAAAGATGACAGCACTGGGTCTATCCGAGAAGCTAAACCAAGAGTGAAGAATA 1839
Mouse: 1381 TGTCTGCCAGTGAGATGAGAAATATTCGATTATCTGACTTCACAGAACTCTTAAAGAA 1440
|||||
Human: 1840 TGTCTGCCAGTGAGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAGAAAA 1899
Mouse: 1441 TAAACCGCAGTGAGTCCCTCAGACCTTAGAAGCATACATACGCTGGAACAAGGATTTG 1500
|||||
Human: 1900 TAAACCGCAGCTGAGCCCTCAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTG 1959
Mouse: 1501 GAGACACCACTGTTTAAAGGAAT 1523
|||||
Human: 1960 GAGATACCACTGTTTAAAGGAAT 1982
Human: 1983 [...] 3263
Mouse: 1524 GGATGCTCTGTGAGCCCATAGAACATCGCACTTCACAGGAAACAAGAGCTTTGGCTACA 1583
1584 GGAACCCAGACTTCGTTTACAGGACGTTTATAGAGTTTTCATTTTGTGCAACCAACTTGA 1643
1644 AGAGGAACAGGAAGACAGACCTAAATAAATATGCAATATGAATGG 1689

FIGURE 6.B

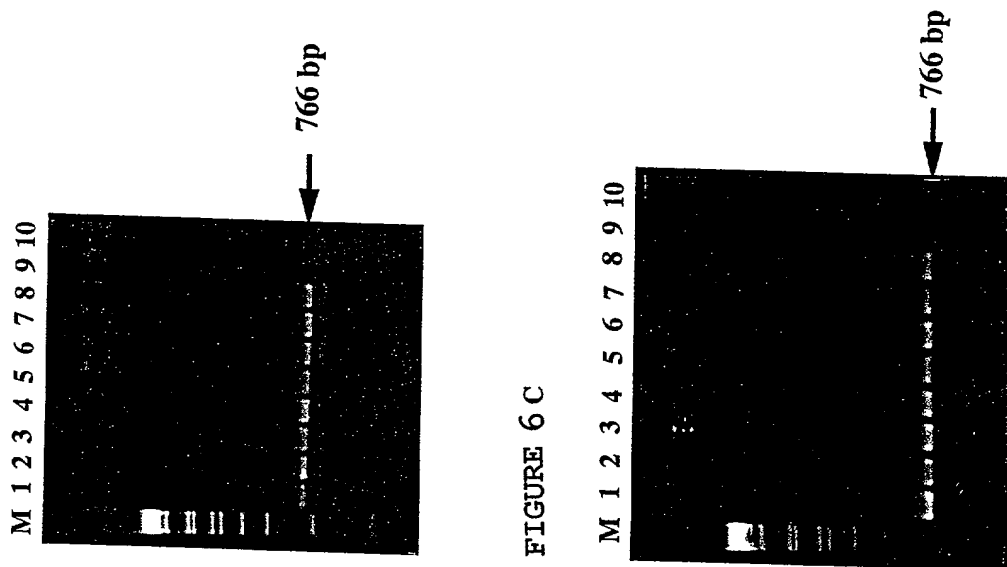
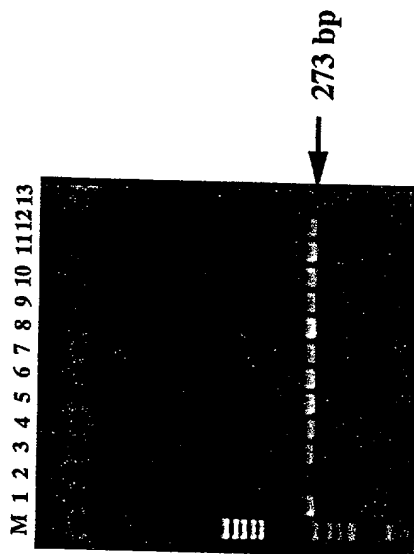


FIGURE 6 A



CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE
RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL DOMINANT
SPASTIC PARAPLEGIA.

5 The invention relates to the identification and characterization of the SPG4 gene
encoding spastin, which is responsible for the most common form of autosomal
dominant hereditary spastic paraplegia (HSP), to the cloning and characterization of its
cDNA, and also to the corresponding polypeptides. The invention also relates to
vectors, to transformed cells and to transgenic animals, and also to diagnostic methods
10 and kits and to methods for selecting a chemical or biochemical compound capable of
interacting directly or indirectly with a polypeptide according to the invention.

Hereditary spastic paraplegias (HSPs) are degenerative disorders of the central
nervous system, characterized by bilateral and progressive spasticity of the lower limbs.
They reveal themselves clinically through difficulties in walking possibly evolving into
15 total paralysis of both legs. The physiopathology of this set of diseases is, to date,
relatively undocumented; however, anatomopathological data make it possible to
conclude that the attack is limited to the pyramidal tracts responsible for voluntary
motricity in the spinal cord (Reid, 1997). Various clinical and genetic forms of HSP
exist. The so-called "pure" HSPs, which correspond to isolated spasticity of the lower
20 limbs, are clinically distinguished from the "complex" HSPs, for which the spasticity of
the legs is associated with other clinical signs of neurological or non-neurological type
(Bruyn et al., 1991). From a genetic point of view, the HSPs can be transmitted
according to the autosomal dominant (AD-HSP), autosomal recessive (AR-HSP) or X-
linked (X-HSP) mode. The "pure" form of HSP, which is most commonly transmitted
25 according to the autosomal dominant mode, remains the most frequent (approximately
80% of HSPs) (Reid, 1997). The incidence of HSPs, which remains difficult to estimate
because of rare epidemiological studies and the considerable clinical variability, varies
from 0.9 : 100 000 in Denmark, 3 to 9.6 : 100 000 in certain regions of Spain (Polo et
al., 1991) or 14 : 100 000 in Norway (Skre, 1974) (approximately 3 : 100 000 in
30 France).

In addition to this great clinical variability, which is observed not only between
various families but also between various affected members of the same family, the
HSPs are also characterized by considerable genetic heterogeneity. In the case of
AD-HSPs, four loci have been identified, to date, on chromosomes 14 (locus SPG3)
35 (Hazan et al., 1993), 2 (locus SPG4) (Hazan et al., 1994; Hentali et al., 1994), 15

(locus SPG6) (Fink et al., 1995) and 8 (locus SPG8) (Hedera et al., 1999). The study of a large number of families exhibiting an AD-HSP has shown that the gene carried by chromosome 2 is a main locus of this form of the disease, found in 40 to 50% of the families analyzed (The Hereditary Spastic Paraplegia Working Group, 1996; Durr et al., 1996). An anticipation phenomenon was observed in some locus SPG4-linked HSP families; this phenomenon has, subsequently, been associated with the expansion of a (CAG)_n repeat demonstrated in 6 Danish families (Nielsen et al., 1997) using the RED (for Rapid Expansion Detection) technique. It has, however, never been possible to confirm this expansion in any of the families tested by this method or by the systematic search for sequences of (CAG)_n type in physical maps composed of YAC (for Yeast Artificial Chromosome) or BAC (for Bacterial Artificial Chromosome) clones (Hazan et al., Genomics, 60 (3), 309-19, 1999).

To date, three genes responsible for two forms of X-HSP and one form of AR-HSP have been identified. Mutations in the gene which encodes a neuron-specific cell adhesion molecule, L1-CAM (for L1 Cell Adhesion Molecule), and which is located at Xq28 (locus SPG1) cause a complex form of HSP (Jouet et al., 1994) in which the spasticity is associated with a mental handicap, whereas mutations in the PLP (for ProteoLipid Protein) gene located at Xq21 (locus SPG2), which encodes a constitutive molecule of the myelin layer, cause pure and complex forms of X-HSP (Saugier-Veber, P. et al., 1994). More recently, mutations in the gene located at 16q24.3 (locus SPG7), which encodes paraplegin, a mitochondrial ATPase of the AAA (for "ATPases Associated with diverse cellular Activities") protein family (Confalonieri et al., 1995), have been associated with complex and pure forms of AR-HSP (Casari et al., 1998).

Thus, there remains, today, a great need to identify and characterize the gene responsible for the most common form of AD-HSP. The identification of this gene should, in particular, allow, besides the possibility of a test for antenatal screening in the families concerned, a better understanding of some of the molecular mechanisms engendering these degenerations specific for nerve bundles of the spinal cord, or even make it possible to provide an elementary response regarding therapeutic treatment for the patients.

This is precisely the subject of the present invention.

After having delimited the localization range between the D2S352 and D2S2347 genetic markers by studying recombination events in locus SPG4-linked HSP families, the inventors have established a contig of BACs covering a physical distance evaluated at approximately 1.5 Mb and have undertaken a positional cloning strategy based on

sequencing the SPG4 range in order to completely identify all the genes located in the candidate region. The analysis of the sequence of the two BACs, D (b336P14) and G (B763N4), has revealed the presence of a gene which is composed of 17 exons, extending over a distance of approximately 100 kb, and which exhibits homology with the genes encoding proteins of the AAA family. Comparison of the sequence of this gene between the healthy and affected individuals of AD-HSP families has made it possible to demonstrate various mutations in the patients.

A subject of the invention is thus the identification and characterization of the SPG4 (or SPAST) gene encoding a novel nuclear member of the AAA family, responsible for the most common form of AD-HSP.

In a first aspect, a subject of the present invention is a purified or isolated nucleic acid of the SPG4 gene, characterized in that it comprises at least 15 consecutive nucleotides, preferably 20, 25, 30, 35, 40, 45, 50, 75, 100 or 200 consecutive nucleotides, of a sequence chosen from the group comprising:

- the sequence SEQ ID No. 1, which is a genomic sequence of the human SPG4 gene;
- the nucleic acid sequences which are homologs or variants of the nucleic acid of sequence SEQ ID No. 1;
- the sequence which is complementary thereto; and
- the sequence of the corresponding RNA thereof.

The present invention relates, of course, to both the DNA and RNA sequences, and also the sequences which hybridize with them, as well as the corresponding double-stranded DNAs.

The terms "nucleic acid", "nucleic acid sequence" or "sequence of nucleic acid", "polynucleotide", "oligonucleotide", "polynucleotide sequence", and "nucleotide sequence", which will be used equally in the present description, will be intended to refer to both a double-stranded DNA, a single-stranded DNA and products of transcription of said DNAs, and/or an RNA fragment, said isolated natural, or synthetic fragments which may or may not include unnatural nucleotides, referring to a precise series of nucleotides, which may or may not be modified, making it possible to define a fragment or a region of a nucleic acid. The expression "natural isolated, or synthetic DNA and/or RNA fragment, which may or may not include unnatural nucleotides" is intended to mean a precise series of nucleotides, which may or may not be modified, making it possible to define a fragment, a segment or a region of a nucleic acid.

It should be understood that the present invention does not relate to the genomic nucleotide sequences in their natural chromosomal environment, i.e. in the

natural state. It involves sequences which have been isolated and/or purified, i.e. they have been removed directly or indirectly, for example by copying, their environment having been at least partially modified.

The term "homologous nucleic acid sequence" is intended to refer to the sequences which have, with respect to the reference nucleic acid sequence, certain modifications, such as in particular a deletion, a truncation, an extension, a chimeric fusion and/or a mutation, in particular a point mutation, and the nucleic acid sequence of which shows at least 80%, preferably 90% or 95%, identity after alignment, with the reference nucleic acid sequence.

For the purpose of the present invention, the term "percentage of identity" between two nucleic acid or amino acid sequences is intended to refer to a percentage of nucleotides or of amino acid residues which are identical between the two sequences to be compared, obtained after the best alignment, this percentage being purely statistical and the differences between the two sequences being distributed randomly and throughout their length. Sequence comparisons between two nucleic acid or amino acid sequences are traditionally carried out by comparing these sequences after having optimally aligned them, said comparison being carried out by segment or by "window of comparison" in order to identify and compare local regions of sequence similarity. The optimal alignment of the sequences for comparison can be produced, besides manually, by means of the local homology algorithm of Smith and Waterman (1981) [Ad. App. Math. 2:482], by means of the local homology algorithm of Needleman and Wunsch (1970) [J. Mol. Biol. 48:443], by means of the similarity search method of Pearson and Lipman (1988) [Proc. Natl. Acad. Sci. USA 85:2444], and by means of computer programs using these algorithms (GAP, BESTFIT, FASTA and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI, or with the BLAST N or BLAST P comparison programs).

The percentage of identity between two nucleic acid or amino acid sequences is determined by comparing these two optimally aligned sequences by window of comparison in which the region of the nucleic acid or amino acid sequence to be compared can comprise additions or deletions with respect to the reference sequence for optimal alignment between these two sequences. The percentage of identity is calculated by determining the number of identical positions for which the nucleotide or the amino acid residue is identical between the two sequences, dividing this number of identical positions by the total number of positions in the window of comparison and

multiplying the result obtained by 100 so as to obtain the percentage of identity between these two sequences.

For example, the BLAST program "BLAST 2 sequences" (Tatusova et al., "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol. Lett. 174:247-250), available on the site <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>, may be used, the parameters used being those given by default (in particular for the parameters "open gap penalty" : 5, and "extension gap penalty" : 2; the matrix chosen being, for example, the "BLOSUM 62" matrix proposed by the program), the percentage of identity between the two sequences to be compared being calculated directly by the program.

It preferably involves sequences for which the complementary sequences are capable of hybridizing specifically with one of the sequences of the invention. Preferably, the specific or high stringency hybridization conditions will be such that they ensure at least 80%, preferably 90% or 95%, identity after alignment between one of the two sequences and the sequence which is complementary to the other.

Hybridization under high stringency conditions means that the temperature and ionic strength conditions are chosen such that they allow the hybridization between two complementary DNA fragments to be maintained. By way of illustration, high stringency conditions of the hybridization step for the purposes of defining the polynucleotide fragments described above are advantageously as follows.

The DNA-DNA or DNA-RNA hybridization is carried out in two steps: (1) prehybridization at 42°C for 3 hours in phosphate buffer (20 mM, pH 7.5) containing 5 x SSC (1 x SSC corresponds to a 0.15 M NaCl + 0.015 M sodium citrate solution), 50% of formamide, 7% of sodium dodecyl sulfate (SDS), 10 x Denhardt's, 5% of dextran sulfate and 1% of salmon sperm DNA; (2) actual hybridization for 20 hours at a temperature dependent on the size of the probe (i.e. 42°C for a probe of size > 100 nucleotides), followed by two 20-minute washes at 20°C in 2 x SSC + 2% SDS and one 20-minute wash at 20°C in 0.1 x SSC + 0.1% SDS. The final wash is carried out in 0.1 x SSC + 0.1% SDS for 30 minutes at 60°C for a probe of size > 100 nucleotides. The high stringency hybridization conditions described above for a polynucleotide of defined size will be adjusted by those skilled in the art for oligonucleotides of greater or smaller size, according to the teaching of Sambrook et al., 1989.

The term "nucleic acid sequence which is a variant" or "nucleic acid which is a variant" of a reference nucleic acid sequence will be intended to refer to the set of nucleic acid sequences corresponding to allelic variants, i.e. individual variations of the

reference nucleic acid sequence. These natural mutated sequences correspond to polymorphisms present in mammals, in particular in human beings, and in particular to polymorphisms which can cause a pathology to occur and/or to develop.

5 While the sequences according to the invention relate to normal sequences, they also relate to sequences which are mutated insofar as they include at least one point mutation, and preferably at most 10% of mutations, with respect to the normal sequence.

10 In particular, the variant nucleic acid sequences will comprise any sequence of at least 15 consecutive nucleotides, preferably 20, 25, 30, 50, 100 or 200 consecutive nucleotides, of a polymorphic sequence of the genomic sequence of the human SPG4 gene of sequence SEQ ID No. 1, and the nucleic acid sequence of which has, with respect to the sequence SEQ ID No. 1, at least one mutation corresponding in particular to a truncation, deletion, substitution and/or addition of an amino acid residue. In the present case, the variant nucleic acid sequences having at least one
15 mutation will herein be linked to the pathologies of AD-HSP type linked to SPG4 locus.

Preferably, the present invention relates to the mutated nucleic acid sequences in which the mutations produce a modification of the amino acid sequence of the polypeptide encoded by the normal sequence.

20 The term "variant nucleic acid sequences" will also be intended to refer to any RNA or cDNA resulting from a mutation of a splice site of the genomic nucleic acid sequence SEQ ID No. 1.

Preferably, the invention relates to a purified or isolated nucleic acid of the SPG4 gene according to the invention, characterized in that it comprises a sequence chosen from the group comprising:

- 25 a) the sequence SEQ ID No. 1, the sequence SEQ ID No. 2, the sequence SEQ ID No. 72, the sequence SEQ ID No. 106 or the sequence of at least 15, preferably 20, 25, 30, 35, 40, 45, 50, 75, 100 or 200, consecutive nucleotides of the sequence SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 or SEQ ID No. 106;
- 30 b) the nucleic acid sequences which are homologs or variants of the sequences SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 or SEQ ID No. 106; and
- c) the complementary sequence or the RNA sequence corresponding to the sequences as defined in a) and b),

preferably with the exception of the nucleic acid identified in the GenBank database under the accession number AB029006.

The nucleic acid the sequence of which is disclosed in the GenBank database under the accession number AB029006 corresponds to the sequence of one of the 100 cDNAs derived from a human brain mRNA library identified by the Kazusa DNA Research Institute in Japan (Kikuno et al., DNA Research, 6, 197-205, 1999).

5 Preferably, the invention relates to a purified or isolated nucleic acid according to the invention, characterized in that it comprises at least one sequence of at least 15 consecutive nucleotides, preferably 20, 25, 30, 50 or 75 consecutive nucleotides, of the nt 714-809, ends inclusive, fragment of the sequence SEQ ID No. 2, of the sequence complementary thereto or of the sequence of the corresponding RNA thereof.

10 The invention preferably relates to a purified or isolated nucleic acid according to the present invention, characterized in that it comprises a sequence chosen from the following group:

- the sequence SEQ ID No. 1;
- the sequence SEQ ID No. 2, which is the cDNA sequence encoding human spastin;
- 15 - the sequences SEQ ID No. 72 and SEQ ID No. 106, the sequence SEQ ID No. 72 representing the sequence of the incomplete cDNA encoding murine spastin represented in Figure 5, "mouse" line, and the SEQ ID No. 106 representing the complete sequence thereof;
- the nucleic acid sequences which are homologs or variants of the sequences SEQ ID
- 20 No. 1, SEQ ID No. 2, SEQ ID No. 72 or SEQ ID No. 106;
- the sequence complementary thereto; and
- the sequence of the corresponding RNA thereof.

25 Preferably, the invention relates to a purified or isolated nucleic acid according to the invention, characterized in that it comprises at least one mutation which corresponds to a natural polymorphism in humans, in particular the position and nature of which are identified in Table 5.

The primers or probes, characterized in that they comprise a sequence of a nucleic acid according to the invention, also form part of the invention.

30 The present invention thus relates to the set of primers which can be deduced from the nucleotide sequences of the invention and which may make it possible to demonstrate said nucleotide sequences of the invention, in particular the mutated sequences, using in particular an amplification method such as the PCR method, or a related method.

35 The present invention also relates to the set of probes which can be deduced from the nucleotide sequences of the invention, in particular from the sequences

capable of hybridizing with them, and which may make it possible to demonstrate said nucleotide sequences, in particular to distinguish the normal sequences from the mutated sequences.

5 The present invention relates, in particular, to the probes or primers having sequences chosen from the sequences SEQ ID No. 4 to SEQ ID No. 71.

The invention also relates to the use of a nucleic acid sequence according to the invention as a probe or primer, for detecting, identifying, assaying or amplifying a nucleic acid sequence.

10 According to the invention, the polynucleotides which can be used as a probe or as a primer in processes for detecting, identifying, assaying or amplifying a nucleic acid sequence will have a minimum size of 15 bases, preferably of 20 bases, or better still of 25 to 30 bases.

15 The set of probes and primers according to the invention may be labeled directly or indirectly with a radioactive or nonradioactive compound, using methods well known to those skilled in the art, in order to obtain a detectable and/or quantifiable signal.

The nonlabeled polynucleotide sequences according to the invention can be used directly as a probe or primer.

20 The sequences are generally labeled so as to obtain sequences which can be used for many applications. The labeling of the primers or of the probes according to the invention is carried out with radioactive elements or with nonradioactive molecules.

25 Among the radioactive isotopes used, mention may be made of ^{32}P , ^{33}P , ^{35}S , ^3H or ^{125}I . The nonradioactive entities are selected from ligands, such as biotin, avidin or streptavidin, dioxxygenin, haptens, colorants and luminescent agents, such as radioluminescent, chemiluminescent, bioluminescent, fluorescent or phosphorescent agents.

30 The polynucleotides according to the invention can thus be used as a primer and/or probe in processes using, in particular, the PCR (polymerase chain reaction) technique (Erlich, 1989; Innis et al., 1990, and Rolfs et al., 1991). This technique requires choosing pairs of oligonucleotide primers framing the fragment which must be amplified. Reference may, for example, be made to the technique described in American patent US No. 4,683,202. The amplified fragments can be identified, for example after agarose or polyacrylamide gel electrophoresis, or after a chromatographic technique such as gel filtration or ion exchange chromatography, and
35 then sequenced. The specificity of amplification can be controlled using, as a primer,

the nucleotide sequences of polynucleotides of the invention and, as a matrix, plasmids containing these sequences or the derived amplification products. The amplified nucleotide fragments can be used as reagents in hybridization reactions in order to demonstrate the presence, in a biological sample, of a target nucleic acid having a sequence complementary to that of said amplified nucleotide fragments.

The invention is also directed toward the nucleic acids which can be obtained by amplification using primers according to the invention.

Other techniques for amplifying the target nucleic acid can be advantageously employed as an alternative to PCR (PCR-like), using pairs of primers having nucleotide sequences according to the invention. The term "PCR-like" will be intended to refer to all methods using direct or indirect reproductions of nucleic acid sequences, or in which the labeling systems have been amplified. These techniques are, of course, known. In general, they involve amplifying the DNA with a polymerase; when the sample of origin is an RNA, it is advisable to perform reverse transcription beforehand. There are, currently, a great many processes which enable this amplification, such as for example the SDA (Strand Displacement Amplification) technique (Walker et al., 1992), the TAS (Transcription-based Amplification System) technique described by Kwoh et al. in 1989, the 3SR (Self-Sustained Sequence Replication) technique described by Guatelli et al. in 1990, the NASBA (Nucleic Acid Sequence Based Amplification) technique described by Kievitis et al. in 1991, the TMA (Transcription Mediated Amplification) technique, the LCR (Ligase Chain Reaction) technique described by Landegren et al. in 1988 and improved by Barany et al. in 1991, which uses a heat-stable ligase, the RCR (Repair Chain Reaction) technique described by Segev in 1992, the CPR (Cycling Probe Reaction) technique described by Duck et al. in 1990, and the Q-beta-replicase amplification technique described by Miele et al. in 1983 and improved, in particular, by Chu et al. in 1986 and Lizardi et al. in 1988, and then by Burg et al., and also by Stone et al., in 1996.

When the target polynucleotide to be detected is an mRNA, use will advantageously be made, prior to carrying out an amplification reaction using the primers according to the invention or carrying out a detection process using the probes of the invention, of an enzyme of reverse transcriptase type in order to obtain a cDNA from the mRNA contained in the biological sample. The cDNA obtained will then serve as a target for the primers or probes used in the amplification or detection process according to the invention.

The probe hybridization technique can be carried out in diverse ways (Matthews et al., 1988). The most general method consists in immobilizing the nucleic acid extracted from the cells of various tissues or from cells in culture, on a support (such as nitrocellulose, nylon or polystyrene), and in incubating the immobilized target nucleic acid with the probe, under well defined conditions. After hybridization, the excess probe is eliminated and the hybrid molecules formed are detected using the appropriate method (measurement of the radioactivity, of the fluorescence or of the enzymatic activity linked to the probe).

According to another embodiment of the nucleic acid probes according to the invention, the latter can be used as a capture probe. In this case, a probe, termed "capture probe", is immobilized on a support and is used to capture, by specific hybridization, the target nucleic acid obtained from the biological sample to be tested, and the target nucleic acid is then detected using a second probe, termed "detection probe", labeled with an easily detectable element.

The splice acceptor or donor site sequences according to the present invention identified in Table 3 (sequences SEQ ID No. 74 to SEQ ID No. 105) also form part of the present invention.

In another aspect, the invention comprises a method for screening cDNA or genomic DNA libraries, or for cloning isolated genomic or cDNA encoding spastin, characterized in that it uses a nucleic acid sequence according to the invention.

Among these methods, mention may be made in particular of :

- the screening of cDNA libraries and the cloning of the isolated cDNAs (Sambrook et al., 1989; Suggs et al., 1981; Woo et al., 1979), using the nucleic acid sequences according to the invention;
- the screening of genomic libraries, for example of BACs (Chumakov et al., 1992; Chumakov et al., 1995), and, optionally, a genetic analysis by FISH (Cherif et al., 1990), using sequences according to the invention, enabling the isolation and chromosomal localization, and then the complete sequencing, of the SPG4 gene encoding spastin.

In particular, these methods according to the invention may be used for identifying and thus obtaining the genomic sequence or the cDNA of the SPG4 gene in other mammals, in particular mice.

These screening and/or cloning methods will comprise, in particular, a step of hybridization of a nucleic acid according to the invention with a nucleic acid contained in a genomic or cDNA library.

The invention also comprises a method for identifying the nucleic acid sequences which promote and/or regulate the expression of the SPG4 gene of sequence SEQ ID No. 1, characterized in that it uses a nucleic acid according to the invention.

5 The computer tools available to those skilled in the art enable them to easily identify, using the genomic nucleic acid sequences according to the invention, the promoter regulatory boxes required and sufficient for controlling gene expression, in particular the TATA, CCAAT and GC boxes, and also the stimulatory regulatory sequences ("enhancers"), or inhibitory regulatory sequences ("silencers"), which
10 control, in CIS, the expression of the genes according to the invention; among these regulatory sequences, mention should be made of IRE, MRE and CRE.

 The invention also relates to the methods for identifying mutations carried by the human SPG4 gene, in particular mutations responsible for autosomal dominant hereditary spastic paraplegia, characterized in that they use a nucleic acid sequence
15 according to the invention.

 These methods for identifying these mutations will, in particular, comprise the following steps: (i) isolation of the DNA from the biological sample to be analyzed, or production of a cDNA from the mRNA of the biological sample; (ii) specific amplification of the target DNA likely to have a mutation, using primers according to the invention;
20 (iii) analysis of the amplification products, in particular the size and/or the sequence of the amplification products, with respect to a reference sequence.

 The expression "methods for identifying a mutation according to the invention" is also intended to refer to a method which makes it possible to obtain the nucleic acid on which said mutation has been identified.

25 The promoter and/or regulatory sequences of the SPG4 gene according to the invention having mutations which may modify the expression of the corresponding protein also form part of the invention.

 The nucleic acids characterized in that they can be obtained using one of the preceding methods according to the invention, or the nucleic acids capable of
30 hybridizing, under high stringency conditions (homology of at least 80% between one of the two sequences and the sequence complementary to the other), with said nucleic acids, form part of the invention, especially the variant or homologous nucleic acids, in particular the nucleic acid sequences of allelic variants of the SPG4 gene of sequence SEQ ID No. 1 or of its cDNA of sequence SEQ ID No. 2, and also the genomic
35 sequences of the homologous genes of other mammals such as mice.

In the present description, the term "Spg4" will be intended to refer to the mouse gene homologous to the human SPG4 gene.

The use of a nucleic acid sequence according to the invention as a probe or primer for screening a genomic library or a cDNA of course forms part of the subject of the present invention.

In another aspect, the invention comprises a purified or isolated polypeptide encoded by a nucleic acid according to the invention, preferably with the exception of the 584 amino acid peptide, the sequence of which is identified in the GenBank database under the accession number AB029006.

In the present description, the term "polypeptide" will be used to refer equally to a protein or a peptide.

Preferably, the present invention relates to a polypeptide according to the invention, characterized in that it comprises an amino acid sequence chosen from the following group:

- the sequence SEQ ID No. 3, corresponding to human spastin encoded by the sequence SEQ ID No. 2 of the cDNA of the human SPG4 gene;
- the sequence SEQ ID No. 73, corresponding to a fragment of murine spastin encoded by the sequence SEQ ID No. 72 of the incomplete cDNA of the mouse Spg4 gene, the sequence SEQ ID No. 73 is represented in Figure 4A, "SPAST_MOUSE" line;
- the sequence SEQ ID No. 107, corresponding to murine spastin encoded by the sequence SEQ ID No. 106 of the complete cDNA of the mouse Spg4 gene;
- the sequences of polypeptides which are homologs and variants of the polypeptide of sequence SEQ ID No. 3, SEQ ID No. 73 or SEQ ID No. 107; and
- the sequences of the fragments thereof of at least 8, 10, 15, 30 or 50 consecutive amino acids.

Also preferably, a subject of the invention is a polypeptide according to the invention, characterized in that it comprises an amino acid sequence chosen from the group comprising:

- a) the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107 or the sequence of at least 10 consecutive amino acids of one of these sequences; and
- b) the sequences which are homologs or variants of the sequences SEQ ID No. 3, SEQ ID No. 73 or SEQ ID No. 107.

Also preferably, a subject of the invention is a polypeptide according to the invention, characterized in that it comprises the sequence of at least 8, preferably of at

least 10, 15, 20 or 30, consecutive amino acids of the sequence of the aa 197-228, ends inclusive, fragment of the sequence SEQ ID No. 3.

Also preferably, a subject of the invention is a polypeptide according to the invention, characterized in that it comprises an amino acid sequence chosen from the following group:

- the sequence SEQ ID No. 3, the sequence SEQ ID No. 73 and the sequence SEQ ID No. 107, which sequences carrying at least one of the mutations corresponding to a natural polymorphism in humans, in particular those the nature and location of which are identified in Table 5 hereinafter, or those which may be identified using the methods for identifying mutations of the SPG4 gene, according to the present invention; and
- the sequences of the fragments thereof of at least 8, 10, 15, 30 or 50 consecutive amino acids.

It should be understood that the invention does not relate to polypeptides in natural form, i.e. they are not taken in their environment. Specifically, the invention relates to the peptides which are obtained by purification from natural sources, or obtained by genetic recombination or by chemical synthesis, and which can therefore include unnatural amino acids. The production of a recombinant polypeptide, which can be carried out using one of the nucleotide sequences according to the invention, is particularly advantageous since it makes it possible to obtain an increased degree of purity of the desired polypeptide.

The term "homologous polypeptide" will be intended to refer to the polypeptides which have certain modifications with respect to the reference polypeptide, such as in particular one or more deletions or truncations, an extension, a chimeric fusion and/or one or more substitutions, and the amino acid sequence of which shows at least 80%, preferably 90% or 95%, identity after alignment, with the reference amino acid sequence.

The term "variant polypeptide" (or protein variant) will be intended to refer to the set of polypeptides encoded by the variant nucleic acid sequences as defined above.

In particular, the variant polypeptides will comprise any polypeptide which is encoded by the mutated genomic sequence of the SPG4 gene of sequence SEQ ID No. 1, and the amino acid sequence of which has at least one mutation corresponding in particular to a truncation, deletion, substitution and/or addition of amino acid residues with respect to the sequence SEQ ID No. 3. In the present case, the variant

polypeptides having at least one mutation will be linked to the pathologies of AD-HSP type.

The term "variant polypeptide" will also be intended to refer to any polypeptide resulting from mutation of a splice site in the genomic nucleic acid sequence SEQ ID No. 1.

The invention also comprises the cloning and/or expression vectors containing a nucleic acid sequence according to the invention.

The vectors according to the invention, characterized in that they include the elements which allow the expression and/or the secretion of said sequences in a host cell, or a cellular addressing sequence, also form part of the invention.

The vectors characterized in that they include a promoter and/or regulator sequence according to the invention also form part of the invention.

Said vectors will preferably include a promoter, translation initiation and termination signals, and also suitable regions for regulating the transcription. They should be able to be maintained stably in the cell and can, optionally, have particular signals which specify secretion of the translated protein.

These various control signals are chosen as a function of the host cell used. To this effect, the nucleic acid sequences according to the invention can be inserted into vectors which replicate autonomously in the host chosen, or vectors which integrate in the host chosen.

Among the systems which replicate autonomously, use will preferably be made, as a function of the host cell, of the systems of plasmid or viral type, the viral vectors possibly in particular being adenoviruses (Perricaudet et al., 1992), retroviruses, lentiviruses, poxviruses or herpesviruses (Epstein et al., 1992). Those skilled in the art know the technology which can be used for each of these systems.

When integration of the sequence into the chromosomes of the host cell is desired, use may be made, for example, of the systems of plasmid or viral type; such viruses will, for example, be retroviruses (Temin, 1986), or AAVs (Carter, 1993).

Among the nonviral vectors, preference is given to naked polynucleotides such as naked DNA or naked RNA according to the technique developed by the company VICAL, yeast artificial chromosomes (YAC) for expression in yeast, mouse artificial chromosomes (MAC) for expression in murine cells and, preferably, human artificial chromosomes (HAC) for expression in human cells.

Such vectors will be prepared according to the methods commonly used by those skilled in the art, and the clones resulting therefrom can be introduced into a

suitable host using standard methods, such as for example lipofection, electroporation or heat shock.

The invention also comprises the host cells, in particular the eukaryotic and prokaryotic cells, transformed with the vectors according to the invention, and also the transgenic animals, except humans, comprising one of said transformed cells according to the invention.

Among the cells which can be used for these purposes, mention may of course be made of bacterial cells (Olins and Lee, 1993), but also yeast cells (Buckholz, 1993), as well as animal cells, in particular cultures of mammalian cells (Edwards and Aruffo, 1993), and especially Chinese hamster ovary (CHO) cells, but also insect cells in which it is possible to use processes implementing baculoviruses, for example (Luckow, 1993). A preferred cellular host for expressing the proteins of the invention consists of CHO cells.

Among the mammals according to the invention, preference will be given to animals such as mice, rats or rabbits, expressing a polypeptide according to the invention.

Among the mammals according to the invention, preference will also be given to those comprising a transformed cell characterized in that the sequence of at least one of the two alleles of the SPG4 gene contains at least one of the mutations corresponding to a natural polymorphism in humans, in particular those the nature and location of which are identified in Table 5 hereinafter, or those which may be identified using the methods for identifying a mutation of the SPG4 gene, according to the present invention.

Among the mammals according to the invention, preference will also be given to animals such as mice, rats or rabbits, characterized in that the gene encoding spastin according to the invention is not functional or is knocked out.

Among the animal models more particularly advantageous herein, there are, in particular:

- the transgenic animals having, at least in one of their two allelic sequences of the SPG4 gene, at least one of the mutations the position and nature of which are identified in Table 5 or identified using a method according to the present invention. These transgenic animals are obtained, for example, by homologous recombination on embryonic stem cells, transfer of these stem cells to embryos, selection of the chimeras affected in the reproductive lines, and growth of said chimeras;

- the transgenic animals (preferably mice) overexpressing the SPG4 gene into which one of said mutations according to the invention may be introduced. The mice are obtained, for example, by transfection of a copy of this gene under the control of a strong promoter which is ubiquitous in nature or selective for a tissue type, or after viral transcription;
- the transgenic animals (preferably mice) made deficient for the SPG4 gene according to the invention by inactivation using the LOXP/CRE recombinase system (Rohlmann et al., 1996) or any other system for inactivating the expression of this gene.

The cells and mammals according to the invention can be used in a method for producing a polypeptide according to the invention, as described below, and can also be used as a model for analysis and for DNA (genomic or cDNA) library screening.

The transformed cells or mammals as described above can thus be used as models in order to study the interactions between the polypeptides according to the invention, and chemical or protein compounds, which are involved directly or indirectly in the activities of the polypeptides according to the invention, this being in order to study the various mechanisms and interactions which come into play.

They can especially be used for selecting products which interact with the polypeptides according to the invention, in particular human spastin of sequence SEQ ID No. 3 or the variants thereof according to the invention, as a cofactor or as an inhibitor, in particular a competitive inhibitor, or which have agonist or antagonist activity for the activity of the polypeptides according to the invention. Preferably, said transformed cells or transgenic animals will be used as a model which, in particular, enables the selection of products which make it possible to combat the pathology linked to the SPG4 gene mentioned above.

The invention also relates to the use of a cell, of a mammal or of a polypeptide according to the invention for screening a chemical or biochemical compound which can interact directly or indirectly with the polypeptides according to the invention, and/or which is capable of modulating the expression or the activity of these polypeptides.

The invention also relates to the use of a nucleic acid sequence according to the invention for synthesizing recombinant polypeptides.

The method for producing a polypeptide of the invention in recombinant form is, itself, included in the present invention, and is characterized in that the transformed cells, in particular the cells or mammals of the present invention, are cultured under conditions which allow the expression of a recombinant polypeptide encoded by a

nucleic acid sequence according to the invention, and in that said recombinant polypeptide is recovered.

The recombinant polypeptides, characterized in that they can be obtained using said production method, also form part of the invention.

5 The recombinant polypeptides obtained as indicated above can be in both glycosylated and nonglycosylated form and may or may not have the natural tertiary structure.

10 These polypeptides can be produced based on the nucleic acid sequences defined above, according to the techniques for producing recombinant polypeptides known to those skilled in the art. In this case, the nucleic acid sequence used is placed under the control of signals which allow its expression in a cellular host.

An effective system for producing a recombinant polypeptide requires a vector and a host cell according to the invention.

15 These cells can be obtained by introducing into host cells a nucleotide sequences inserted into a vector as defined above, and then culturing said cells under conditions which allow the replication and/or expression of the transfected nucleotide sequence.

20 The processes for purifying a recombinant polypeptide which are used are known to those skilled in the art. The recombinant polypeptide can be purified from cell lysates and extracts and/or from the culture medium supernatant, with methods used individually or in combination, such as fractionation, chromatography methods, immunoaffinity techniques using specific monoclonal or polyclonal antibodies, etc.

25 The polypeptides according to the present invention can be obtained by chemical synthesis, this using one of the many known peptide syntheses, for example the techniques which implement solid phases or techniques which use partial solid phases, by condensation of fragments or by conventional synthesis in solution.

The solid-phase synthesis technique is well known to those skilled in the art. See in particular Stewart et al. (1984) and Bodansky (1984).

30 The polypeptides which are obtained by chemical synthesis and which can include corresponding unnatural amino acids are also included in the invention.

The mono- or polyclonal antibodies or their fragments, chimeric antibodies or immunoconjugates, characterized in that they are capable of specifically recognizing a polypeptide according to the invention, form part of the invention.

35 Specific polyclonal antibodies can be obtained from a serum of an animal immunized against the polypeptides according to the invention, in particular produced

by genetic recombination or by peptide synthesis, according to conventional procedures.

The advantage of antibodies which specifically recognize certain polypeptides, variants or immunogenic fragments thereof, according to the invention, will in particular
5 be noted.

The specific monoclonal antibodies can be obtained according to the conventional hybridoma culture method described by Köhler and Milstein, 1975.

The antibodies according to the invention are, for example, chimeric antibodies, humanized antibodies, or Fab or F(ab')₂ fragments. They can also be in the form of
10 labeled antibodies or immunoconjugates in order to obtain a detectable and/or quantifiable signal.

The invention also relates to methods for detecting and/or purifying a polypeptide according to the invention, characterized in that they use an antibody according to the invention.

The invention also comprises purified polypeptides, characterized in that they
15 are obtained using a method according to the invention.

Moreover, besides their use for purifying the polypeptides, the antibodies of the invention, in particular the monoclonal antibodies, can also be used for detecting these polypeptides in a biological sample.

They thus constitute a means of immunocytochemically or immuno-
20 histochemically analyzing the expression of the polypeptides according to the invention, in particular the polypeptide of sequence SEQ ID No. 3 or a variant thereof, on specific tissue sections, for example by immunofluorescence or gold labeling, or with an enzymatic immunoconjugates.

They may make it possible, in particular, to demonstrate abnormal expression
25 of these polypeptides in the biological samples or tissues, which makes them useful for monitoring the progression of the disease and the molecular diagnosis.

More generally, the antibodies of the invention can be advantageously used in any situation in which the expression of a normal or mutated polypeptide according to
30 the invention must be observed.

The methods for determining allelic variability, a mutation, a deletion, a loss of heterozygosity or any genetic abnormality of the SPG4 gene, according to the invention, characterized in that they use a nucleic acid sequence or an antibody according to the invention, also form part of the invention.

The present invention thus comprises a method for genotypic diagnosis of the pathology associated with the SPG4 gene, characterized in that a nucleic acid sequence according to the invention is used.

5 Preferably, the invention relates to a method for genotypic diagnosis of the disease associated with the presence of at least one mutation on a sequence of the SPG4 gene, using a biological sample from a patient, characterized in that it includes the following steps:

- a) where appropriate, isolation of the genomic DNA from the biological sample to be analyzed, or production of cDNA from the RNA of the biological sample;
- 10 b) specific amplification of said DNA sequence of the SPG4 gene likely to contain a mutation, using primers according to the invention;
- c) analysis of the amplification products obtained and comparison of their sequence with the corresponding normal sequence of the SPG4 gene.

15 The invention also comprises a method for diagnosing the disease associated with abnormal expression of a polypeptide encoded by the SPG4 gene, in particular the polypeptide of sequence SEQ ID No. 3, characterized in that one or more antibodies according to the invention is (are) brought into contact with the biological material to be tested, under conditions which allow the possible formation of specific immunological complexes between said polypeptide and said antibody or antibodies, and in that the
20 immunological complexes possibly formed are detected and/or quantified.

These methods are, for example, directed toward the methods for diagnosis, in particular antenatal diagnosis, of AD-HSP associated with the presence of a mutation in the SPG4 gene, according to the invention, by determining, using a biological sample from the patient, the presence of mutations in at least one of the sequences
25 described above. The nucleic acid sequences analyzed may equally be genomic DNA, cDNA or mRNA.

Nucleic acids or antibodies based on the present invention may also be used to enable positive diagnosis in a patient or presymptomatic diagnosis in an individual at risk, in particular an individual with a family history of the disease.

30 There are, of course, a great number of methods which make it possible to demonstrate a mutation in a gene with respect to the wild-type gene. They can essentially be divided into two main categories. The first type of method is that in which the presence of a mutation is detected by comparing the mutated sequence with the corresponding wild-type sequence, and the second type is that in which the presence

of the mutation is detected indirectly, for example through evidence of mismatches due to the presence of the mutation.

These methods can use the probes and primers of the present invention which have been described. They are generally purified nucleic acid hybridization sequences comprising at least 15 nucleotides, preferably 20, 25 or 30 nucleotides, characterized in that they can hybridize specifically with a nucleic acid sequence according to the invention.

Preferably, the specific hybridization conditions are such as those defined above or in the examples. The length of these nucleic acid hybridization sequences can range from 15, 20 or 30 to 200 nucleotides, particularly from 20 to 50 nucleotides.

Among the methods for determining allelic variability, a mutation, a deletion, a loss of heterozygosity or a genetic abnormality, preference is given to the methods comprising at least one so-called PCR (polymerase chain reaction) or PCR-like amplification step for the target sequence according to the invention likely to have an abnormality, using a pair of primers having nucleotide sequences according to the invention. The amplified products may be treated with a suitable restriction enzyme before carrying out the detection and assaying of the product targeted.

The mutations of the SPG4 gene according to the invention may be responsible for various modifications of the translation product thereof, these modifications possibly being used for a diagnostic approach. Specifically, the antigenicity modifications linked to these mutations may allow the development of specific antibodies. The mutated gene product can be distinguished using these methods. All these modifications can be employed in a diagnostic approach, using several well-known methods based on the use of mono- or polyclonal antibodies which recognize the normal polypeptide or mutated variants, such as for example by RIA or by ELISA.

Thus, a subject of the invention is also a kit or pack for diagnosis, in particular for diagnosing AD-HSP associated with the presence of a mutation in the SPG4 gene, according to the invention, characterized in that it comprises at least one compound chosen from the following group of compounds:

- a) a nucleic acid, in particular as a primer or probe, according to the present invention;
- and
- b) an antibody according to the invention.

In another aspect, the invention comprises a method for selecting a chemical or biochemical compound capable of preventing and/or treating AD-HSP associated with the SPG4 gene, characterized in that a nucleic acid sequence according to the

invention, a polypeptide according to the invention, a vector according to the invention, a cell according to the invention, a mammal according to the invention or an antibody according to the invention is used.

The methods for selecting chemical or biochemical compounds capable of interacting directly or indirectly with polypeptides according to the invention or with the nucleic acids according to the invention, and/or making it possible to modulate the expression or the activity of these polypeptides, characterized in that they comprise bringing a polypeptide according to the invention, a transformed cell according to the invention or a mammal according to the invention into contact with a candidate compound, and detecting a modification of the activity of said polypeptide, are also included in the invention.

For example, but without being limited thereto, mention may be made of a method for identifying molecules capable of interacting with a polypeptide according to the invention, using a bacterial or yeast two hybrid system such as the Matchmaker Two Hybrid System 2, according to the instructions of the manual which is supplied with the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech).

The nucleic acids encoding proteins which interact with the promoter and/or regulatory sequences of the SPG4 gene, according to the invention, can be screened and/or selected using a one hybrid system such as that described in the manual which is supplied with the Matchmaker One Hybrid System kit from Clontech (Catalog No. K1603-).

In other aspect, the invention comprises the use of a nucleic acid or of a polypeptide according to the invention, of a vector according to the invention, of a cell according to the invention or of a mammal according to the invention, for studying the expression or the activity of the SPG4 gene.

Other characteristics and advantages of the invention appear in the remainder of the description with the examples and figures, the legends of which are given hereinafter.

LEGENDS OF THE FIGURES

FIGURES 1A, 1B and 1C : Physical map of the SPG4 range and genomic organization of SPG4.

FIGURE 1A : The 1.5 Mb candidate region is delimited by the D2S352 and D2S2347 genetic markers indicated in bold characters. The position of the polymorphic markers and other STSs is indicated in standard characters, whereas the position of

the ESTs is indicated in italics. The BAC clones constituting the presequencing map are represented by rectangles, with the name shown above and the precise size of the clone, if it could be determined, shown below. The name of the BACs A, B, C, etc. is followed by brackets containing the name of the clone preceded by a "b" if the clone is derived from the BACs library CITB_978_SKB, or by a "B" if it originates from the library RPCI-11.

FIGURE 1B : Schematic representation of the SPG4 gene which overlaps BACs D (b336P14) and G (B563N4). The exons are shown as black rectangles with their name above.

FIGURE 1C : The five mutations identified in seven SPG4 locus-linked AD-HSP families are positioned in exons 7, 11 and 13 and in the splice acceptor site of intron 15.

FIGURE 2 : Nucleic acid and protein sequence of the SPG4 cDNA of spastin.

The 17 vertical bars with a number located below represent the junctions between the various exons. The ATG initiator codon is located at nt position 126-128 and the STOP codon for termination is located at nt position 1974-1976. Five of the mutations identified to date, including the loss of exon 16, are indicated in italics (nt 1210, nt 1468, nt 1520, nt 1620 and for the loss of exon 16: nt 1813-1853). The polyadenylation site is in italics and underlined. The putative nuclear localization signal (NLS), RGKKK, and also the three conserved domains predicted by the analysis in the ProDom database are located at aa positions 7-11 (NLS), 342-409 (domain 92), 411-509 (domain 179) and 512-599 (domain 6226), respectively. The four motifs predicted by the sequence comparison in the Prosite database are: two "leucine zipper" motifs at aa positions 50-78 and 508-529, the ATP binding site (or Walker A motif) at aa positions 382-389 and the "helix-loop-helix" dimerization domain at aa positions 478-486. The Walker A and B motifs, "GPPGNGKT" and "IIFIDE", and also the AAA minimum consensus [lacuna] are underlined.

FIGURES 3A and 3B : Characterization of a splice site mutation in the affected individuals of three SPG4 locus-linked AD-HPS families.

FIGURE 3A : PCR amplification of fragment IV of the SPG4 cDNA using lymphoblast cDNA: well M, size marker VII (Boehringer); well 1, unaffected member of family 2992; well 2, patient of family 2992; well 3, unaffected member of family 5330; well 4, patient of family 5330; well 5, patient of family 5226; well 6, negative control (human genomic DNA).

FIGURE 3B : Sequence graph for the mutation of the splice acceptor site of intron 15.

Genomic sequence of the control individual above and of a patient of family 2992 below. The asterisk at nt position 1813-4 indicates an A->C polymorphism which affects a nonconserved nucleotide of the splice acceptor site of intron 15 in the patient. FIGURES 4A and 4B : Spastin homologies.

The identical residues are highlighted by shaded areas.

FIGURE 4A : Multiple alignment created by CLUSTAL W of eight proteins derived from various organisms and having strong sequence homology with human spastin and murine spastin (SEQ ID No. 73).

FIGURE 4B : Alignment by CLUSTAL W of the yeast metalloproteases AFG3, RCA1 and YME1, and of human plaraplegin and spastin.

FIGURE 5: Alignment by BLASTN of the nucleic acid sequences of the SPG4 cDNA and of its mouse ortholog Spg4 (SEQ ID No. 72). The polyadenylation site of the murine cDNA is underlined and in italics. The STOP codon is located at nt position 1515-1517 in the murine cDNA and at nt position 1974-1976 in the human cDNA.

FIGURES 6A, 6B and 6C : PCR analysis of the expression of SPG4 and of its murine ortholog Spg4.

FIGURE 6A : Collection of cDNA originating from multiple mouse tissues.

Well M, size marker V (Boehringer); well 1, heart; well 2, brain; well 3, spleen; well 4, lung; well 5, liver; well 6, skeletal muscle; well 7, kidney; well 8, testicle; well 9, E7 7-day embryo; well 10, E11 11-day embryo; well 11, E15 15-day embryo; well 12, E17 17-day embryo; well 13, negative control (mouse genomic DNA).

FIGURE 6B : Collection of cDNA originating from multiple human tissues.

Well M, size marker VII (Boehringer); well 1, brain; well 2, heart; well 3, kidney; well 4, liver; well 5, lung; well 6, pancreas; well 7, placenta; well 8, skeletal muscle; well 9, negative control (human genomic DNA); well 10, negative control (no DNA).

FIGURE 6C : Collection of cDNA originating from multiple human fetal tissues.

Well M, size marker VII (Boehringer); well 1, brain; well 2, heart; well 3, kidney; well 4, liver; well 5, lung; well 6, skeletal muscle; well 7, spleen; well 8, thymus; well 9, negative control (human genomic DNA); well 10, negative control (no DNA).

EXAMPLES

Example 1: Materials and methods

1) Subcloning and sequencing of the candidate region

Twelve BACs originating from two human genomic libraries, CITB_978_SKB (sold by Research Genetics) and RPCI-11 (Osoegawa et al., 1998), and covering the SPG4 range, were selected to be sequenced (Hazan et al., Genomics, 60 (3), 309-19, 1999). 40 µg of the DNA of each BAC were partially digested with the CviJI restriction enzyme (CHIMERx) and separated by electrophoresis on 0.4% LMP agarose gel (FMC). DNA fractions, the sizes of which vary in the region of 3, 5 and 10 kb, were eluted with β-agarase (Biolabs) and ligated to a plasmid vector pBAM3, which had been digested with SmaI and dephosphorylated, beforehand, in a ratio of 1 × insert per 5 × vector. Electrocompetent *E. coli* DH10B bacteria (GIBCO-BRL) were transformed with the various ligations, by electroporation. Approximately 1 000 to 1 500 subclones per BAC (8 to 10 equivalent genomes), consisting of 20% of clones with inserts at 10 kb, 40% of clones with inserts at 5 kb and 40% of clones with inserts at 3 kb, were isolated. The ends of the inserts of these clones were sequenced on a LICOR 4200 automatic sequencer. For each BAC, the sequences were assembled into a backbone consisting of several contigs, using the Phred and Phrap programs. The holes between each contig were sequenced with labeled dideoxynucleotides on an ABI 377 sequencer (PE-Applied Biosystems). The exons contained in these sequence contigs were predicted with the GRAIL II, GENSCAN, FGENEH and Genie computer programs. The sequences were also compared in the EMBL and GenBank nucleic acid and protein databases, with the BLASTN and BLASTX programs. The determination of the promoter sequences was carried out using the TSSG and TSSW computer programs. The results of all these sequence analyses were visualized using the Genotator sequence annotation program.

2) cDNA cloning

The cDNA of the SPG4 gene was isolated through 5' and 3' RACE-PCR experiments on polyA⁺ RNAs of fetal brain, adult brain and adult liver, using the Marathon cDNA amplification kit (Clontech) according to the supplier's instructions. A first PCR followed by an internal PCR were carried out with various pairs of primers, the sequences of which are indicated in Table 1 hereinafter:

Table 1
Primers used for the RACE-PCRs and the cDNA amplifications

Primer	Sequence (5'-3')	5' position pair/PCR product size			
SPA_5RACE5	CGGAGCTCCTCTTGGCTGCCATG (SEQ ID No.4)	nt 405			
SPA_5RACE6	AGAAGCGCTGGCAGAGCCACACGAAG (SEQ ID No.5)	nt 372			
SPA_5RACE7	AAGGCGACCAAACGCAGCAGCGCGAAG (SEQ ID No.6)	nt 331			
SPA_3RACE1	AGGAGCAAGCTGTGGAATGGTATAAG (SEQ ID No.7)	nt 550			
SPA_3RACE2	TGGTTATGGCCAAGGACCGCTTACAAC (SEQ ID No.8)	nt 689			
SPA_3RACE3	CAAACGGACGTCTATAATGACAGTAC (SEQ ID No.9)	nt 747			
SPA_3RACE4	TTAGGAATGTGGACAGCAACCTTGC (SEQ ID No.10)	nt 1075			
SPA_3RACE5	CTTCTCTGAGGCCTGAGTTGTTCAC (SEQ ID No.11)	nt 1207			
SPA_3RACE6	TGCTAGAATGACTGATGGATACTCAGG (SEQ ID No.12)	nt 1736			
SPA_3RACE7	AGATGCAGCACTGGGTCCTATCCG (SEQ ID No.13)	nt 1787			
SPA_3RACE8	ATGAACGTCATCGGCTACAGAAACAG (SEQ ID No.14)	nt 2037			
SPA_Db	TAGCAGTGGCTGCCGCCGT (SEQ ID No.15)	nt 45	b+m	655 bp	
SPA_Dm	AAGCGGTCCTTGGCCATAAC (SEQ ID No.16)	nt 700			
SPA_Dc	GGCGGCAGTGAGAGCTGTG (SEQ ID No.17)	nt 106	c+n	543 bp	
SPA_Dn	CTAGCTCTTTCACACTGTTC (SEQ ID No.18)	nt 649			
SPA_Ad	AACAGGCCTTCGAGTACATC (SEQ ID No.19)	nt 487	d+n	746 bp	
SPA_Am	CTGTGAACAACCTCAGGCCTC (SEQ ID No.20)	nt 1233			
SPA_Ac	ATGAGAAAGCAGGACAGAAG (SEQ ID No.21)	nt 532			
SPA_An	TGCCAAGTCTTGACCAGC (SEQ ID No.22)	nt 1175			
SPA_Ba	CTACAACTGCTACTCGTAAG (SEQ ID No.23)	nt 1036	a+m	763 bp	
SPA_Bm	CAGTGCTGCATCTTTTGCC (SEQ ID No.24)	nt 1799			
SPA_Bb	TAGGAATGTGGACAGCAACC (SEQ ID No.25)	nt 1076			
SPA_Bn	AAAGCTGTTAGGTCACCTCC (SEQ ID No.26)	nt 1780			
SPA_Ca	TGGAGATGACAGAGTACTTG (SEQ ID No.27)	nt 1550	a+m	766 bp	
SPA_Cm	CTGGAATACTTTCATCTGC (SEQ ID No.28)	nt 2316			
SPA_Cb	ATGAGGCTGTTCTCAGGCG (SEQ ID No.29)	nt 1603			

The RACE-PCR products were cloned with the TA-cloning kit (Invitrogen) and the corresponding clones were sequenced on an ABI 377 (PE-Applied Biosystems). The sequence of the SPG4 transcript was verified by sequencing PCR products amplified from a cDNA population originating from the lymphoblasts of 6 healthy individuals.

3) Detection of mutations

The total RNAs were extracted from lymphoblast lines of one affected individual per family studied and of 6 control individuals, using the RNA PLUS kit (bioprobe System). The cDNA synthesis was carried out on 500 ng to 1 µg of RNA, with 100 pmol of random hexameric primers (Pharmacia) and 200 units of Superscript II reverse transcriptase (Gibco BRL), under standard conditions. Four PCR amplifications, generating overlapping fragments which cover all of the SPG4 open reading frame, were carried out on the cDNAs of the patients and controls. Fragment I was amplified with the SPA_Db/SPA_Dm primers, and then by internal PCR with the SPA_Dc/SPA_Dn primers. Fragments II, III, and IV were amplified with the SPA_Ad/SPA_Am, SPA_Ba/SPA_Bm and SPA_Ca/SPA_Cm primers (cf. the sequences of these primers in Table 1), respectively. Each amplification was carried out in a total volume of 50 µl containing 4 µl of cDNA (~ 1/7th of the prep.), 20 pmol of each primer, 200 µM of dNTPs, 50 mM of KCl, 10 mM of Tris, pH 9, 1.5 mM MgCl₂, 0.1% of triton X-100, 0.01% of gelatin and 2.5 units of Taq polymerase (Cetus-PE). The PCR reactions were carried out according to the "hot start" process: the Taq polymerase is added at 92°C, after a first denaturation step of 5 min at 94°C. The samples are subsequently subjected to 35 cycles of denaturation (94°C for 40 sec), of hybridization (55°C for 50 sec, with the exception of fragment I: 58°C for 50 sec) and of elongation (72°C for 1 min), followed by a final elongation step (5 min at 72°C). The PCR products are sequenced on an ABI 377 automatic sequencer (PE-Applied Biosystems), with the SPA_Dc/SPA_Dn, SPA_Ac/SPA_An, SPA_Bb/SPA_Bn and SPA_Cb/SPA_Cm primers for fragments I, II, III and IV, respectively.

The mutations were also sought or confirmed by sequencing the 17 predicted exons of the SPG4 gene in the patients and controls. Each exon was amplified with the corresponding "a+m" pair of primers (cf. Table 2 hereinafter), with the exception of exon 1 (gSPAex1c/gSPAex1m), and exons 10, 11 and 12 which were co-amplified with the gSPAex10a/gSPAex12m and gSPAex11a/gSPAex12m pairs of primers.

Table 2
PCR primers for amplifying and sequencing the exons

Exon	Product size	PCR program	Primer	Sequence (5'-3') (SEQ ID Nos.; 30 to 71)
1	1048 bp	0	gSPAex1c	GTGAGCCGAACTGCACATTG
			gSPAex1m	CAAAGTCGACAGCTACAGTGC
			gSPAex1d	GGAAGTGTAGTTGAGTGGGA
			gSPAex1n	AGATGAGGCTCCGACCTAC
2	624 bp	3	gSPAex2a	AATGCCACACTTGTAATCTC
			gSPAex2m	TGTGAATATATCATAATTTGGG
			gSPAex2b	TACAGCAGTTCTCATGATG
3	812 bp	1	gSPAex3a	GACCAAATTGGTGCATGCATG
			gSPAex3m	ACATTTCCAATACATCCCAC
4	379 bp	3	gSPAex4a	ATTTGTCATTTACATGCAC
			gSPAex4m	TTAGAATGACTATACCTGAC
			gSPAex4n	TCAGGTAAAGTAAGACTC
5	830 bp	4	gSPAex5a	TTCCTATCTACCTAGTGAC
			gSPAex5m	TTTTATAGCAAGTTGCCCTG
			gSPAex5b	CCTATGAAGATCCTGGTAC
6	484 bp	3	gSPAex6a	TGTCATGATTCTAACAAGGG
			gSPAex6m	TCTATTTCACTCCTGACATG
7	420 bp	2	gSPAex7a	GTCATAGGGCTTAGGCTTC
			gSPAex7m	ATCATACTACCCACTTTTCC
8	647 bp	3	gSPAex8a	TGTTTGGGAAGATGCTACTG
			gSPAex8m	CTACTGAAGATAACGTACATG
9	1268 bp	1	gSPAex9a	CATTGATTGCCATGTATTGG
			gSPAex9m	AGAAGGCCAGAAATACTCAG
			gSPAex9b	GTACTIONAATCGGTAAATATGG
10]	1061 bp	4	gSPAex10a	CTCAAGTCTTAGGAATGCAG
11			gSPAex10b	GCACTTAACCAGGCTGTATG
12]	551 bp	3	gSPAex11a	CTCAGATGACTCACATAGC
			gSPAex12m	CTTTACTAGACTAATTCTCCTG

13	1361 bp	4	gSPAex13a	CAGATTCAAGAAGACAGATC
			gSPAex13m	GCAATAATTCACCACACTTG
			gSPAex13n	GGTAGTTCTTGTTTCTGCTC
14	985 bp	4	gSPAex14a	CAAGTGTGGTGAATTATTGC
			gSPAex14m	GAGCTGAAAAGTATTCAGC
			gSPAex14n	TGCAAAGGACATAGCCAGTG
15	1076 bp	1	gSPAex15a	AGCCTCTGGAGATAGTATGC
			gSPAex15m	CTAGAACAGGGGTCACAGTC
			gSPAex15n	TTGGACTTCTTAACTTC
16	1404 bp	4	gSPAex16a	GCAGTATGCAAGAAATTGAAC
			gSPAex16m	GGCCTGTAATTTTCTTCTG
			gSPAex16b	GTACTGAATAGATACATGTAG
17	445 bp	3	gSPAex17a	GTGTAGCAGATCAACATAG
			gSPAex17m	CATCTTCAAGTTTGGTGCAC

Other than for exon 1, which is amplified using the Advantage GC genomic PCR kit (Clontech) according to the supplier's instructions, four slightly different PCR programs (1, 2, 3 and 4) were used to amplify the SPG4 exons (see Table 2). The amplifications were all carried out in a volume of 50 μ l containing 100 ng of genomic DNA, 50 pmol of each primer, 250 μ M pf dNTPs, 1X Takara buffer and 1 unit of Takara La Taq Taq polymerase (Shuzo Co.). The PCR reactions were carried out according to the "hot start" process: the Taq polymerase is added at 94°C, after a first denaturation step of 5 min at 96°C. The samples are subsequently subjected to 30 cycles of denaturation (94°C for 40 sec), of hybridization (prog. 1: 60°C for 50 sec; prog. 2: 58°C for 50 sec, prog. 3 and 4: 55°C for 50 sec) and of elongation (prog. 1 and 4: 72°C for 1 min, prog. 2 and 3: 72°C for 40 sec), followed by a final elongation step (10 min at 72°C). The sequencing of these PCR products was carried out on an ABI 377 sequencer (PE-Applied Biosystems), using either the PCR primers or the internal primers termed "b" and "n" (see Table 2).

4) Characterization of SPG4

The cDNA clones 977312 (EST AA560327) and 568234 (EST AA107866) derived from the mouse blastocyst and E8 embryo cDNA libraries, which both correspond to the murine ortholog of SPG4, were isolated using the IMAGE consortium

and sequenced in the laboratory on an ABI 377 sequencer (PE-Applied Biosystems). In order to analyze the expression profile of SPG4 and of its murine ortholog Spg4, the collections of cDNA from various fetal and adult human tissues, and also from mouse tissues (MTC panels, Clontech), were tested by PCR according to the supplier's protocol, with the SPA_Ca/SPA_Cm pair of primers for the human cDNAs and the SPA_Ca/spam (spam: 5'-ACCGAAGTCAAGAGCCTATC-3') pair for the mouse cDNAs. The PCR conditions are those used for amplifying SPG4 from lymphoblast line cDNA (cf. § Detection of mutations), except that these samples were subjected to 32 cycles for the cDNAs derived from adult human tissues and from mouse tissues, and to 28 cycles for the cDNAs derived from fetal tissues. The amplification products migrated by electrophoresis on 2% agarose gels.

5) Histological analysis of a muscle biopsy from a patient

The histological and histo-enzymatic analyses were carried out on a muscle biopsy from a patient derived from an SPG4 locus-linked family according to the standard techniques described in Casari et al., 1998.

6) Accession numbers in the public databases

The SPG4 (or SPAST) cDNA and the deduced protein sequence, GenBank/EMBL AJ246001; the incomplete Spg4 cDNA clone, GenBank/EMBL AJ246002; the SPG4 (or SPAST) gene, GenBank/EMBL AJ246003.

Example 2 : Analysis of the sequence of the SPG4 range

The analysis of the recombination events made it possible to reduce the SPG4 candidate region to a genetic range of 0 cM between the D2S352 and D2S2347 markers (19, 20). A presequencing map of the SPG4 range composed of 37 BACs was constructed (Hazan et al., in press in Genomics); the candidate region covers a physical distance of approximately of 1.5 Mb. Twelve overlapping BACs, stretching over the SPG4 region, with the exception of a single 4 kb hole between clones A and E, were selected to be sequenced (fig. 1A). Seven of these BACs (A, B, C, D, E, F and G), covering approximately 70% of the region of interest, have already been sequenced. The sequences of these 7 BACs were compared with those of the nucleic acid and protein databases, and analyzed with four exon prediction programs. These preliminary sequence analyses made it possible to reveal 14 potential transcription units, including three corresponding to the genes encoding xanthine dehydrogenase, steroid 5 α -reductase 2 and a TGF β -binding protein. Of the 14 genes detected by the sequence analysis, 9 had been previously identified in the EST (for "Expressed Sequence Tag") databases and located in the SPG4 range (Hazan et al., in press in

Genomics); the 5 remaining genes could only be identified by sequencing the candidate region. One of these 5 novel genes showed homology in 3' of its coding region, with the genes encoding the AAA protein family (Confalonieri et al., 1995). More thorough sequence analyses showed that this gene, named SPG4 (or SPAST), was composed of 17 exons and extended over a region of approximately 90 kb, covered by two adjacent BAC clones, D and G (cf. fig. 1B). The first three predicted exons of this gene were identified in BAC D, by two of the four exon prediction programs used, GRAIL II and GENSCAN; they show strong homology with a mouse blastocyst EST, AA560327. The last 14 exons are found in BAC G. The protein sequence deduced from exons 7 to 17 is significantly homologous to a subclass of the AAA family, which includes the Yta6p (Schnall et al., 1994), TBP6 (Schnall et al., 1994) and End 13 yeast proteins, and also the SKD1 mouse protein (Perier et al., 1994).

Of the four exon prediction programs FGENEH appears to be the most reliable and the most powerful, enabling detection of most of the genes of this chromosomal region at 2p21-p22. This observation also applies to the SPG4 gene, for which 15 exons could be demonstrated using this program, while only 4, 9 or 11 exons could be located using the Genie, GRAIL II and GENSCAN programs, respectively. The genomic organization of this gene (fig. 1B) could subsequently be confirmed by determining the sequence of the SPG4 cDNA. The intron/exon junctions are represented on table 3 hereinafter: the exon size ranges from 41 bp (exon 16) to 1.410 kb (exon 17), that of the introns ranging from 140 bp (intron 11) to 23.247 kb (intron 1).

Table 3

Intron/exon organization of the SPG4 gene

Exon/ intron	Exon size (bp)	Position on the cDNA	Splice acceptor site (SEQ ID No. 74 to 89)	Splice donor site (SEQ ID Nos. 90 to 105)	Intron size (bp)
1	540	1		TGAGAAAG/gtaactaggggctgg	23 247
2	87	541	atttttattttaaag/CAGGACAG	AGGACAAAG/gtaagattgtattgt	1 943
3	84	628	aatTTTTctttcag/GTGAACAG	ACTTCTAG/gtatcaattaatgtat	9 190
4	96	712	cttctcgtgcatag/AGAAAGATG	CCAGTCAG/gtgggttaggttaac	15 745
5	188	808	acttttccctgtcag/AAAGTGGA	CTCATAAG/gtattctgggacagta	876
6	134	996	tttggatccctttaag/GGTACTCC	GTGGACAA/gtaagtttggccatct	283
7	94	1 130	aggctctgtttcttag/TGGAACAG	GGCCTGAG/gtaagaaccttatatt	10 735
8	75	1 224	agtatataTTTTtag/TTGTTTAC	CAATGCTG/gtaaggggtctcttca	1 385
9	72	1 299	cttggaattttaaag/GCTAAAGC	CAAAATAC/gtgagtgctctgttcc	8 083
10	76	1 371	taatgcttggTTTTtag/GTGGGAGA	TTTTATAG/gtaagaacacatatttc	238
11	92	1 447	cttgatttctctctag/ATGAAAGTT	TTGATGGT/gtaagtggtgattatg	140
12	80	1 539	gaTTTTtgctgttag/GTACAGTC	GTTCTCAG/gtagggagatttatat	4 715
13	43	1 619	ggattTTTTtttag/GCGTTTCA	ATGAGGAG/gtatgtatctgtgttt	1 389
14	80	1 662	ttttaataattttcag/ACAAGACT	CTTGCTAG/gtgagtaatttgatt	1 521
15	71	1 742	tccttcccttctcag/AATGACTG	TATCCGAG/gtaggtatacaagagc	2 210
16	41	1 813	cttttatgttttacag/AACTAAAA	CCAGTGAG/gtatagtattttacaa	7 115
17	1 410	1 854	cttttaaaaaatctag/ATGAGAAA		

The sequences of the exons and introns are indicated in upper case and lower case, respectively.

Example 3 : Identification of the SPG4 cDNA

Several successive amplifications by 5' and 3' RACE-PCR were carried out on collections of adult liver and brain and fetal brain cDNA, in order to characterize the SPG4 transcript. All the 5' RACE-PCRs gave amplification products terminating at nt position 263 of the SPG4 cDNA (fig. 2), which was probably due to the rich GC content of the 5' region of the transcript (90% of GC in the 60 bp preceding nt position 263). Four overlapping PCR products, covering all of the coding region, were amplified from the cDNAs derived from the lymphoblasts of six control individuals, and entirely sequenced with the aim of verifying the sequence of the SPG4 transcript. Aligning the sequences of all the PCR and RACE-PCR products made it possible to reconstitute a 3263 bp sequence comprising a 1848 bp open reading frame preceded by a 125 bp untranslated 5' region (5' UTR for "5' UnTranslated Region") and followed by 1290 bp 3' UTR region including a polyadenylation site between nt positions 3227-3232, ~ 35 bp upstream of the polyA tail (fig. 2). Comparing the sequence of the SPG4 cDNA with the EST databanks made it possible to detect significant homology with 6 human ESTs, including EST N47973 which contains a more extended 3' noncoding region (+ 180 bp) comprising a second polyadenylation site. The translation initiation site was identified by the presence of a Kozak consensus sequence (CTGTGAatgA) defined as a "suitable context" for translation initiation given that a purine is located 3 nt upstream of the initiator ATG, itself preceded by a STOP codon. The 3263 bp cDNA sequence is identical to the transcribed sequence deduced from the 17 exons of the SPG4 gene. The analysis of the sequence of the 5' region using the TSSG and TSSW computer programs suggests the presence of a promoter sequence of the TATA box type located 43 bp upstream of nt position 1 of exon 1.

Example 4 : Mutations in the SPG4 gene

Heterozygous mutations were sought in the SPG4 cDNA originating from lymphoblasts of 14 patients derived from SPG4 locus-linked families (1 affected individual per family). Four overlapping PCR fragments, I, II, III and IV, covering the open reading frame of the SPG4 cDNA, were amplified and sequenced in the 14 patients, and also in 6 healthy control individuals. The agarose gel electrophoresis of PCR fragment IV showed three bands of equal intensity in 3 patients from families 2992, 5226 and 5330 originating from the same region of Switzerland, which would suggest a microdeletion or a mutation of a splice site; the two additional bands were not present in 2 healthy individuals derived from families 2992 and 5330 (fig. 3A). The genomic sequence of exon 16 revealed a heterozygous A->G mutation of the splice acceptor site (AG) of intron 15 in the affected

individuals of these three families (fig. 3B); this mutation engenders the loss of exon 16, followed by a reading frame shift in the abnormal transcript. None of the healthy members, including husbands and wives, carry this mutation of the splice site. The identification of the same mutation in all the affected members of these three Swiss families demonstrates the existence of a common ancestor, which had probably been suggested by the study of the haplotypes.

Three point mutations, 1210C->G, 1468G->A and 1620C->T, which introduced amino acid substitutions into the protein sequence (S362C, C448Y and R499C), were respectively revealed by sequencing PCR fragments III and IV in the affected individuals of families 624, 4014 and 618. These three substitutions all involve a cysteine residue, inducing the loss or insertion of a cysteine in the protein sequence. A 1 bp deletion, 1520delT, which creates the appearance of a STOP codon inducing a truncated protein composed of 465 amino acids (aa), was detected in the affected individuals of family A. None of the five mutations summarized in table 4 hereinafter was found in the control individuals tested, whether they belong to the healthy siblings or to the spouses of the seven families analyzed herein. These five mutations significantly affect the protein sequence in a very conserved domain, or AAA cassette (Beyer, 1997), which is composed of several protein motifs presumed to be responsible for the ATPase activity in all the members of the AAA family.

Table 4
Mutations in SPG4 in the patients suffering from AD-HSP

Family	Location	Mutation ^a	Amino acid change ^b	Consequence
624	exon 7	1 210 C → G	S362C	missense
4 014	exon 11	1 468 G → A	C448Y	missense
A	exon 11	1 520 delT	466STOPcodon	nonsense
618	exon 13	1 620 C → T	R499C	missense
2 992	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5 226	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5 330	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift

^a The nt positions refer to the sequence of the SPG4 cDNA.

^b The aa positions refer to the spastin sequence.

The bases of the exons are indicated in upper case, those of the introns in lower case.

PTC+7 aa = "premature termination codon" at 7 aa downstream of exon 16.

In addition to these five mutations described above, searches for heterozygous mutations, carried out on patients suffering from AD-HSP derived from 36 other families, made it possible to reveal 34 other mutations which modified or were likely to modify the product of expression of the SPG4 gene.

- 5 The characteristics of these 34 other mutations are summarized in table 5 hereinafter, into which the first five mutations mentioned above have also been inserted.

Table 5
Mutations in SPG4 in the patients suffering from AD-HSP

Family	Location	Mutation ^a	Amino acid change ^b	Consequence
624	exon 7	1210 C → G	S362C	missense
6958	exon 8	1233 G → A	G370R	missense
214	exon 8	1267 T → G	F381C	missense
1002	exon 8	1283 T → G	N386K	missense
027	exon 8	1288 A → G	K388R	missense
019	exon 10	1401 C → G	L426V	missense
4014	exon 11	1468 G → A	C448Y	missense
148	exon 11	1504 G → T	R460L	missense
618	exon 13	1620 C → T	R499C	missense
636	exon 15	1788 G → A	D555N	missense
627	exon 15	1792 C → T	A556V	missense
2971	exon 3	702 C → T	Q193STOP	nonsense
3655	exon 5	873 A → T	K229STOP	nonsense
1010	exon 5	907 C → A	S261STOP	nonsense
3938	exon 5	932 C → G	Y269STOP	nonsense
6922	exon 10	1416 C → T	R431STOP	nonsense
616	exon 10	1416 C → T	R431STOP	nonsense
605	exon 15	1809 C → T	R562STOP	nonsense
030	exon 2	578-579insA	PTC + 2 aa	shift + nonsense
615	exon 5	852del11	PTC + 18 aa	shift + nonsense
042	exon 5	882-883insA	PTC + 12 aa	shift + nonsense
032	exon 5	906delT	PTC + 17 aa	shift + nonsense
189	exon 9	1299delG	PTC + 3 aa	shift + nonsense
3686	exon 9	1340del5	PTC + 35 aa	shift + nonsense
625	exon 9	1340del5	PTC + 35 aa	shift + nonsense
A	exon 11	1520delT	PTC + 7 aa	shift + nonsense
115	exon 12	1574delGG	PTC + 2 aa	shift + nonsense
3266	exon 13	1634del22	PTC + 18 aa	shift + nonsense
149	exon 14	1684-1685insTT	PTC + 9 aa	shift + nonsense
645	exon 14	1685del4	PTC + 7 aa	shift + nonsense
029	intron 4	808-2 a → g	?	splice site mutation
162	intron 6	1129+2 t → g	?	splice site mutation
125	intron 7	1223+1 g → t	?	splice site mutation
143	intron 8	1299+1 g → a	(PTC + 6 aa)	splice site mutation
1620	intron 11	1538+5 g → a	?	loss of exon 11 + shift
1006	intron 11	1538+3 del4	?	splice site mutation
1605	intron 13	1661+1 g → t	?	splice site mutation
1012	intron 13	1662-2 a → t	?	splice site mutation
1626	intron 15	1812+1 g → a	Δ aa564 → aa576 (PTC+7 aa)	splice site mutation
2992	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5226	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5330	intron 15	1813-2 a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
1611	intron 16	1813-2 a → g	?	splice site mutation
		1853+1 g → a		

^a The nt positions refer to the sequence of the SPG4 cDNA. ^b The aa positions refer to the spastin sequence. The exon bases are indicated in upper case, those of the introns in lower case. PTC+n aa - "premature termination codon" at n amino acids downstream of the mutation.

Example 5 : Analysis of the protein sequence of spastin

The open reading frame of SPG4 encodes a 616 aa protein which we have named spastin and the molecular weight of which is approximately 67.2 kDaltons (kD). The comparison of this amino acid sequence in the protein databases, using the BLAST programs, made it possible to reveal a region of strong homology with several members of the AAA family, at the C-terminal end of spastin. The "typical" motifs of the AAA family, encompassed in the AAA cassette, are located between aa positions 342 and 599 (see fig. 2) according to the sequence comparisons in the ProDom and Prosite protein domain databases. The three conserved typical domains, including the Walker A and B motifs and also the minimum consensus motif of the AAA proteins are located in the AAA cassette at aa positions 382-389, 437-442 and 480-498, respectively, (fig. 2). The Walker A motif, "GPPGNGKT", also called p-loop, which corresponds to the ATP-binding domain, and the B motif, "IIFIDE", are very conserved among all the members of the AAA family, including spastin.

The comparison of the AAA cassettes present in 150 proteins of this ATPase family, derived from organisms which are very far apart in evolution made it possible to classify this set of proteins into several subgroups, as a function of the number of AAA cassettes identified (1 or 2) and of the sequence homologies between these various cassettes (Beyer, 1997). Among all the proteins of the AAA family, spastin shows stronger homology with a particular subclass of the AAAs, and more specifically with the following proteins, most of which were identified through the complete sequencing of the genome of the organism in question: two proteins of *Caenorhabditis elegans*, O16299 and Q18128; two subunits of the 26S proteasome of *Saccharomyces cerevisiae*, Yta6p (Q02845) and TBP6 (P40328) (Schnall et al., 1994); a subunit of the proteasome of *Schizosaccharomyces pombe* (O43078); the SAP1 (P39955) and END13 (P52917) proteins of *S. cerevisiae* and the murine SKD1 protein (P46467) (Perier et al., 1994). The multiple alignment of these 8 proteins with spastin is represented in fig. 4A. Of the 257 amino acids encompassing the AAA cassette (aa positions 342-599), spastin shows 52%, 51% and 50% sequence identity with the Yta6p (Q02845) yeast protein, the O16299 nematode protein and the TBP6 (P40328) yeast protein, respectively. Similar results were obtained by analyzing the protein sequence of spastin in the ProDom database, which showed the existence of three domains of homology (named 92, 179 and 6226, and corresponding to aa positions 342-409, 411-509 and 512-599) found in the putative subunits of the 26S proteasome of yeast. In addition, the members of this AAA subgroup most commonly contain motifs of the leucine-zipper type, two of which could be detected

in the protein sequence of spastin at aa positions 50-78 and 508-529, by analyzing the sequence in the Prosite database (see fig. 2). This analysis was also able to predict the presence of a dimerization motif of the helix-loop-helix type, located between aa positions 478 and 486.

5 The comparison of the protein sequence of spastin with those of the mitochondrial metalloproteases, such as the AFG3, RCA1 and YME1 yeast proteins, and also paraplegin, which is implicated in a rare form of AR-HSP, shows that the homology between these five members of the AAA family is limited to the 257aa region encompassing the AAA cassette (fig. 4B). In this region, the sequence identity between
10 spastin and paraplegin is only 29%, whereas paraplegin and the AFG3 yeast protein are 57% identical over this same portion of the protein sequence. This sequence comparison suggests that spastin does not belong to the same AAA subgroup as paraplegin and other mitochondrial metalloproteases. In addition, the computer analysis of the spastin sequence using the PSORT II program, which makes it possible to predict the subcellular
15 location of the proteins, appears to indicate that spastin is a nuclear protein. A possible nuclear localization signal (NLS), RGKKK, was revealed between aa positions 7 and 11, whereas no signal peptide characteristic of importation into mitochondria could be detected, unlike what had been observed for paraplegin.

Example 6 : Expression profiles for SPG4 and for its murine ortholog Spg4

20 The comparison of the nucleic acid sequence of SPG4 in the EST databanks made it possible to detect several human, murine and rat ESTs showing strong homology with SPG4. The mouse blastocyst and E8 embryo cDNA clones corresponding to two of the murine ESTs, AA560327 and AA107866, were obtained from the IMAGE consortium and entirely sequenced. The assembly of the sequences of these cDNA clones made it
25 possible to reconstitute a 1689 bp consensus sequence including a 1514 bp incomplete open reading frame. The comparison between the human SPG4 cDNA and this mouse cDNA showed that the murine transcript lacks approximately 460 bp at the 5' end, including the translation initiation codon. The mouse open reading frame is followed by a 175 bp 3' noncoding region (3' UTR) containing a polyadenylation site located ~20 bp
30 upstream of the polyA tail (fig. 5). The nucleic acid sequence of SPG4 and the protein sequence of human spastin show 89% (between nt positions 460 and 1982) and 96% (between aa positions 113 and 616) identity, respectively, with the mouse cDNA and deduced protein sequences. This considerable degree of homology makes it possible to affirm that this mouse transcript corresponds to the murine ortholog of SPG4, which was
35 therefore named Spg4.

The hybridization of Northern blots comprising the mRNAs of various human and murine tissues (Clontech) with the SPG4 and Spg4 cDNA clones did not give any convincing results, except a very weak band corresponding to a 2.5 kb transcript in the mouse testicle after exposure for 10 days. Because of the low level of expression of this gene, the expression profiles for SPG4 and Spg4 were determined by PCR experiments on normalized collections of cDNA originating from various adult and fetal tissues (see fig. 6A to 6C). The murine Spg4 gene is expressed ubiquitously in the adult tissues of mice, and also from the E7 stage to the E17 stage of mouse embryos (fig. 6A). Higher expression of Spg4 was detected in the liver, skeletal muscle and testicles, and also at the E15 stage of embryos. The early expression of Spg4 during embryonic development was confirmed by the presence of ESTs originating from blastocyst, E8 embryo and embryonic carcinoma cDNA libraries in the public EST databanks. The human SPG4 gene is, itself, also expressed ubiquitously in adult (fig. 6B) and fetal (fig. 6C) tissues, with perhaps more marked expression in fetal brain.

Example 7 : No oxidative phosphorylation impairment in SPG4 locus-linked AD-HSP

In order to determine whether spastin mutations induced an oxidative phosphorylation (OXPHOS) impairment in mitochondria, in the same way as had been observed for paraplegin, a muscle biopsy was performed on a patient from one of the SPG4 locus-linked AD-HSP families. The morphological and histo-enzymatic analyses of this muscle biopsy did not reveal any muscle fibers of the RRF (for "ragged red fiber") type, characteristic of OXPHOS impairments in mitochondria. The fact that all the muscle fibers appear to be normal, and also the prediction of a nuclear localization for spastin, seem to indicate that SPG4 locus-linked AD-HSP is not a mitochondrial disease of the OXPHOS type, unlike SPG7 locus-linked AR-HSP.

Using a positional cloning approach based on sequencing a 1.5 Mb region, we have identified the SPG4 (or SPAST) gene responsible for the most common form of AD-HSP, previously located on chromosomal bands 2p21-p22. Thirty nine mutations which modify or are likely to modify the gene product, named spastin, could be detected in the affected individuals from forty one families with AD-HSP showing a link to the SPG4 locus. Spastin is a novel member of the AAA protein family, which appears to have a nuclear localization and which shows strong homology with the subunits of the 26S proteasome of yeast. Despite great homology restricted to a domain of 230 to 250 aa, termed AAA cassette, the many members of this protein family can participate in very varied cellular mechanisms, such as the transport of proteins in vesicles, cell cycle

regulation, organelle biogenesis, i.e. control of transcription, etc. However, all these cellular mechanisms involve the assembly, the functioning or the degradation of protein complexes, which suggest that the members of the AAA family are so-called "chaperon" proteins.

References

- Barany, F., (1991), *Proc. Natl. Acad. Sci. USA*, 88, 189-193.
- Beyer, A. Sequence analysis of the AAA protein family. *Protein Sci.* 6, 2043-2058 (1997).
- Bodansky M., *Principles of peptide synthesis*, (1984).
- 5 Bruyn, R.P.M. & Scheltens, P.H. Hereditary spastic paraparesis (Strumpell-Lorrain) in *Handbook of clinical neurology* Vol. 15 (ed. de Jong, J.M.B.V.) 301-318 (Elsevier Science Publishers B.V., 1991).
- Buckholz, R.G. *Curr. Op. Biotechnology* 4 : 538-542, 1993.
- Burg, J.L. et al. (1996), *Mol. and Cell. Probes*, 10, 257-271.
- 10 Carter, B.J. *Curr. Op. Biotechnology* 3 : 533-539, 1993.
- Casari, G. et al. Spastic paraplegia and OXPHOS impairment caused by mutations in Paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* 93, 973-983 (1998).
- Cherif D., Julier, C., Delattre, O., Derré, J., Lathrop, G.M., and Berger, R. *Proc. Natl. Acad. Sci. USA*. 87 : 6639-6643, 1990.
- 15 Chu, B.C.F. et al. (1986), *Nucleic Acids Res.*, 14, 5591-5603.
- Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billault, A., Guasconi, G., Gervy, P., Le Chumakov, I.M., Rigault, P., Le Gall, I., et al. *Nature* 377 : 175-183, 1995.
- Confalonieri, F. & Duguet, M. A 200-amino acid ATPase module in search of a basic function. *BioEssays* 17, 639-650 (1995).
- 20 Duck, P. et al. (1990), *Biotechniques*, 9, 142-147.
- Durr, A. et al. Phenotype of autosomal spastic paraplegia linked to chromosome 2. *Brain* 119, 1487-1496 (1996).
- Edwards, C.P., and Aruffo, A. *Curr. Op. Biotechnology* 4 : 558-563, 1993.
- Epstein, A. *Médecine/Sciences* 8 : 902-911, 1992.
- 25 Erlich, H.A., (1989), New York : Stockton Press.
- Guatelli J.C. et al. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878, 1990 et al. *Cell* 85 : 281-290, 1996.
- Fink, J.K. et al. Autosomal dominant familial spastic paraplegia : tight linkage to chromosome 15q. *Am. J. Hum. Genet.* 56, 188-192 (1995).
- 30 Hazan, J., Lamy, C., Melki, J., Munnich, A., de Recondo, J., & Weissenbach, J. Autosomal dominant familial spastic paraplegia is genetically heterogeneous and one locus maps to chromosome 14q. *Nature Genet.* 5, 163-167 (1993).
- Hazan, J. et al. Linkage of a new locus for autosomal dominant familial spastic paraplegia to chromosome 2p. *Hum. Mol. Genet.* 3, 1569-1573 (1994).

- Hedera, P. et al. Novel locus for autosomal dominant hereditary spastic paraplegia, on chromosome 8q. *Am. J. Hum. Genet.* 64, 563-569 (1999).
- Heinzlef, O. et al. Mapping of a complicated familial spastic paraplegia to locus SPG4 on chromosome 2p. *J. Med. Genet.* 35, 89-93 (1998).
- 5 Hentati, A. et al. Linkage of a locus for autosomal dominant familial spastic paraplegia to chromosome 2p markers. *Hum. Mol. Genet.* 3, 1867-1871 (1994).
- The Hereditary Spastic Paraplegia Working Group. Hereditary spastic paraplegia : advances in genetic research. *Neurology* 46, 1507-1514 (1996).
- Innis, M.A. et al. (1990), Academic Press.
- 10 Jouet, M. et al. X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the L1 gene. *Nature Genet.* 7, 402-407 (1994).
- Kievitis, T. et al. (1991), *J. Virol. Methods*, 35, 273-286.
- Köhler et Milstein. *Nature* 256, 495-497, 1975.
- Kwoh, D.Y. et al. (1989), *Proc. Natl. Acad. Sci. USA*, 86, 1173-1177.
- 15 Landegren U., Kaiser R., Sanders J. & Hood L. *Science* 241 : 1077-1080, 1988.
- Lizardi, P.M. et al. (1988), *Bio/technology*, 6, 1197-1202.
- Luckow, V.A. (1993), *Curr. Op. Biotechnology* 4, 564-572.
- Matthews, J.A. et al. (1988), *Anal. Biochem.*, 169 : 1-25.
- Miele, E.A. et al. (1983), *J. Mol. Biol.*, 171 : 281-295.
- 20 Nielsen, J.E. et al. CAG repeat expansion in autosomal dominant pure spastic paraplegia linked to chromosome 2p21-24. *Hum. Mol. Genet.* 6, 1811-1816 (1997).
- Olins, P.O., and Lee, S.C. *Curr. Op. Biotechnology* 4 : 520-525, 1993.
- Osoegawa, K. et al. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52, 1-8 (1998).
- 25 Perier, F. et al. Identification of a novel mammalian member of the NSF/CDC48p/Pas1p/TBP-1 family through heterologous expression in yeast. *FEBS lett.* 351, 286-290 (1994).
- Perricaudet, M., Stratford-Perricaudet, L. and Briand, P. *La Recherche* 23 : 471-473, 1992.
- 30 Polo, J.M., Calleja, J., Combarros, O. & Berciano, J. Hereditary ataxias and paraplegias in Cantabria, Spain. An epidemiological and clinical study. *Brain* 114, 855-866 (1991).
- Reid, E. Pure hereditary spastic paraplegia. *J. Med. Genet.* 34, 499-503 (1997).
- Rohlmann, A., Gotthardt, M., Willnow, T.E., Hammer, R.E., and Herz, J. *Nature Biotech.* 14 : 1562-1565, 1996.
- 35 Rolfs, A. et al. (1991), Berlin : Springer-Verlag.

- Sambrook, J., Fritsch, E.F., and Maniatis, T. Molecular cloning : a laboratory manual.
- Saugier-Veber, P. et al. X-linked spastic paraplegia and Pelizaeus-Merzbacher disease are allelic disorders at the proteolipid protein locus. *Nature Genet.* 6, 257-262 (1994).
- Schnall, R. et al. Identification of a set of yeast genes coding for a novel family of putative
- 5 ATPases with high similarity to constituents of the 26S protease complex. *Yeast* 10, 1141-1155 (1994).
- Scott, W.K. et al. Locus heterogeneity, anticipation, and reduction of the chromosome 2p minimal candidate region in autosomal dominant familial spastic paraplegia. *Neurogenetics* 1, 95-102 (1997).
- 10 Sec. Ed. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
- Segev, D., (1992), Kessler C. Springer Verlag, Berlin, New-York, 197-205.
- Skre, H. Hereditary spastic paraplegia in Western Norway. *Clin. Genet.* 6, 165-183 (1974).
- Stone, B.B. et al. (1996). *Mol. and Cell. Probes*, 10 : 359-370.
- Stewart J.M. et Yound J.D., solid phase peptides synthesis, Pierce Chem. Company,
- 15 Rockford, 111, 2ème éd., (1984).
- Suggs S.V., Wallace R.B., Hirose T., Kawashima E.H. and Itakura K. *PNAS* 78 : 6613-6617, 1981.
- Temin, H.M. Retrovirus vectors for gene transfer. In Kucherlapati R., ed. *Gene Transfer*, New York, Plenum Press, 149-187, 1986.
- 20 Walker G.T., Fraiser M.S., Schram J.L., Little M.C., Nadeau J.G., & Malinowski D.P. *Nucleic Acids Res.* 20 : 1691-1696, 1992.
- Werderlin, L. Hereditary ataxias. Occurrence and clinical features. *Acta Neurol. Scand.* 73 (Suppl. 106) (1986).
- Woo S.L.C. *Methods Enzymol.* 68 : 389, 1979.

CLAIMS

1. Purified or isolated nucleic acid of the SPG4 gene, characterized in that it comprises a sequence chosen from the group comprising:

- 5 a) the sequence SEQ ID No. 1, the sequence SEQ ID No. 2, the sequence SEQ ID No. 72, the sequence SEQ ID No. 106 or the sequence of at least 15 consecutive nucleotides of one of these sequences;
- b) the nucleic acid sequences which are homologs or variants of the sequences SEQ ID No. 1, SEQ ID No. 2, SEQ ID No. 72 or SEQ ID No. 106; and
- 10 c) the complementary sequence or the RNA sequence corresponding to the sequences as defined in a) and b).

2. Purified or isolated nucleic acid according to claim 1, with the exception of the nucleic acid identified in the GenBank databank under the accession number AB029006.

- 15 3. Purified or isolated nucleic acid according to claim 1 or 2, characterized in that it comprises at least one sequence of at least 15 consecutive nucleotides of the nt 714-809, ends inclusive, fragment of the sequence SEQ ID No. 2, of the sequence complementary thereto or of the sequence of the corresponding RNA thereof.

4. Purified or isolated nucleic acid according to one of claims 1 to 3, characterized in that it comprises a mutation corresponding to a natural polymorphism in humans.

5. Probe or primer, characterized in that it comprises a sequence of a nucleic acid according to one of claims 1 to 4.

6. Probe or primer according to claim 5, characterized in that its sequence is chosen from the sequences SEQ ID No. 4 to SEQ ID No. 71.

7. Splice acceptor or donor site, characterized in that it comprises a sequence of a nucleic acid according to claim 1 chosen from the sequences SEQ ID No. 74 to SEQ ID No. 105.

8. Method for screening cDNA or genomic DNA libraries, or for cloning isolated genomic or cDNA encoding spastin, characterized in that it uses a nucleic acid sequence according to one of claims 1 to 7.

9. Method according to claim 8, for identifying the genomic or cDNA sequence of the SPG4 gene of mammals, in particular of mice.

10. Method for identifying a mutation carried by the human SPG4 gene, characterized in that it uses a nucleic acid sequence according to one of claims 1 to 7.

11. Method according to claim 10, for identifying a mutation responsible for autosomal dominant hereditary spastic paraplegia.

12. Method for identifying the nucleic acid sequences which promote and/or regulate the expression of the SPG4 gene, characterized in that it uses a nucleic acid sequence according to one of claims 1 to 7.

13. Nucleic acid identified using a method according to one of claims 9 to 12.

14. Polypeptide encoded by a nucleic acid according to one of claims 1 to 4 and 13.

15. Polypeptide according to claim 14, preferably with the exception of the 584 amino acid peptide, the sequence of which is identified in the GenBank databank under the accession number AB029006.

16. Polypeptide according to claim 14 or 15, characterized in that it comprises an amino acid sequence chosen from the group comprising:

a) the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107 or the sequence of at least 10 consecutive amino acids of one of these sequences; and

b) the sequences which are homologs or variants of the sequences SEQ ID No. 3, SEQ ID No. 73 or SEQ ID No. 107.

17. Polypeptide according to claim 14 or 15, characterized in that it comprises the sequence of at least 8 consecutive amino acids of the sequence of the aa 197-228, ends inclusive, fragment of the sequence SEQ ID No. 3.

18. Polypeptide according to claim 14 or 15, characterized in that it comprises an amino acid sequence chosen from the group comprising the sequence SEQ ID No. 3, the sequence SEQ ID No. 73, the sequence SEQ ID No. 107, which sequences carrying at least one of the mutations corresponding to a natural polymorphism in humans, and the sequences of the fragments thereof of at least 10 consecutive amino acids.

19. Cloning and/or expression vector containing a nucleic acid sequence according to one of claims 1 to 4, and 13.

20. Vector according to claim 19, characterized in that it includes the elements required for its expression in a host cell.

21. Host cell transformed with a vector according to claim 19 or 20.

22. Mammal, except a human, characterized in that it comprises a cell according to claim 21.

23. Mammal, except a human, according to claim 22, comprising a transformed cell, characterized in that the sequence of at least one of the two alleles of the SPG4 gene

contains at least one of the mutations corresponding to a natural polymorphism in humans or identified using a method according to claim 10 or 11.

24. Use of a nucleic acid sequence according to one of claims 5, 6 and 13, as a probe or primer, for detecting and/or amplifying nucleic acid sequences.

5 25. Use of a nucleic acid sequence according to one of claims 1 to 7, and 13, for screening a genomic or cDNA library.

26. Use of a nucleic acid sequence according to one of claims 1 to 4 and 13, for producing a recombinant or synthetic polypeptide.

10 27. Method for producing a recombinant polypeptide, characterized in that a transformed cell according to claim 21 is cultured under conditions which allow the expression of said recombinant polypeptide, and in that said recombinant polypeptide is recovered.

28. Polypeptide, characterized in that it is obtained using a method according to claim 27.

15 29. Mono- or polyclonal antibodies or their fragments, chimeric antibodies or immunoconjugates, characterized in that they are capable of specifically recognizing a polypeptide according to one of claims 14 to 18, and 28.

30. Method for detecting and/or purifying a polypeptide according to one of claims 14 to 18, and 28, characterized in that it uses an antibody according to claim 29.

20 31. Method for genotypic diagnosis of AD-HSP associated with the SPG4 gene, characterized in that a nucleic acid sequence according to one of claims 1 to 7 and 13 is used.

25 32. Method for genotypic diagnosis of AD-HSP associated with the presence of at least one mutation on a sequence of the SPG4 gene, using a biological sample from a patient, characterized in that it includes the following steps:

- 30 a) where appropriate, isolation of the genomic DNA from the biological sample to be analyzed, or production of cDNA from the RNA of the biological sample;
- b) specific amplification of said DNA sequence of the SPG4 gene likely to contain a mutation, using primers according to either of claims 5 and 6 or a nucleic acid according to claim 13;
- c) analysis of the amplification products obtained and comparison of their sequence with the corresponding normal sequence of the SPG4 gene.

35 33. Method for diagnosing AD-HSP associated with abnormal expression of a polypeptide encoded by the SPG4 gene, characterized in that one or more antibodies according to claim 29 is (are) brought into contact with the biological material to be tested,

34. Method for selecting a chemical or biochemical compound which is capable of interacting directly or indirectly with a polypeptide according to one of claims 14 to 18, and 28, or with a nucleic acid according to one of claims 1 to 7, and 13, and/or which makes it possible to modulate the expression or the activity of these polypeptides, characterized in that it comprises bringing a nucleic acid sequence according to one of claims 1 to 7, and 13, a polypeptide according to one of claims 14 to 18, and 28, a vector according to either of claims 19 and 20, a cell according to claim 21, a mammal according to either of claims 22 and 23 or an antibody according to claim 29 into contact with a candidate compound, and detecting a modification of the activity of said polypeptide.

36. Kit or pack for diagnosis, characterized in that it comprises at least one compound chosen from the following group of compounds:

20 a) a nucleic acid according to either of claims 5 and 6; and

b) an antibody according to claim 29.

1/7

FIGURE 1A

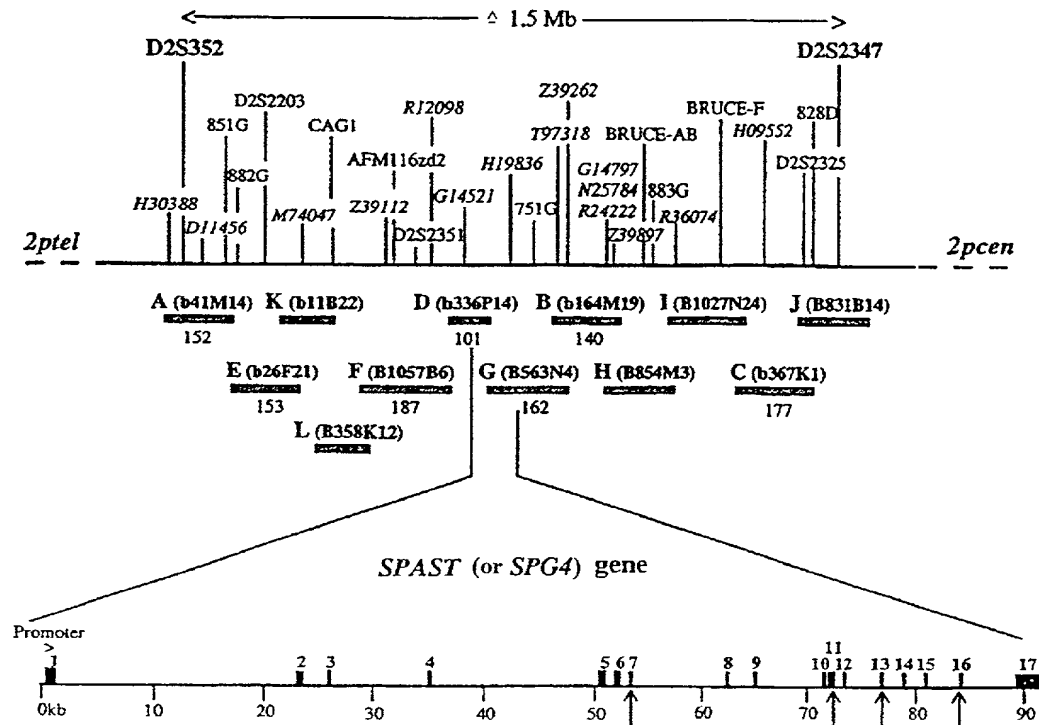
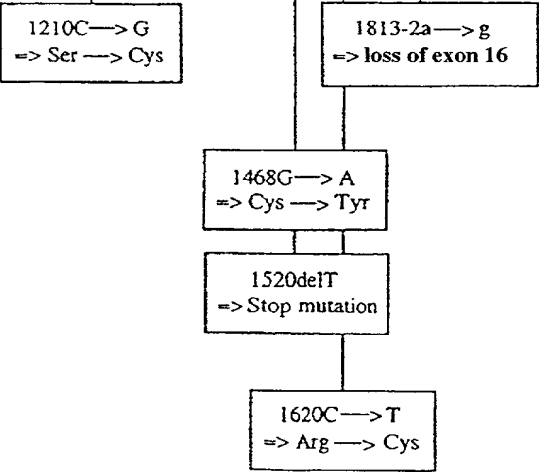


FIGURE 1B

FIGURE 1C



1 | GCTCCTGAGACCGCGCGGCACACGGGGTCTGTGGCCCCCGCTAGCAGTGGCTGCGCGCTGCTTGGTTCCCGTGGTCTCGGGAGCGCGG 95
1 | TTATGCGCGCGCGCGCAGTGAAGCTGTGAATGAATTCCTCGCGTGCACGAGGGAAGAAAGGCTCCGCGCGCGCGCAGCAACCGGTGCTCTCC 190
1 | M N S P G G R G K K K G S G G A S N P V P P
1 | CAGGCTCCGCCCCCTGCTGCCCCCGCCCCCTCCCGCGCGCGCGCGCGCGCTCCGCGCGAGTCCCGCATAAGCGAACTGTACTATTCT 285
23 | R P P P P C L A P A P P A A G P A P P P E S P H K R N L Y Y F S
23 | CCTACCCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTGGCGCTTCCACCTGGGGCTCTCTTCTGTTGGCTCTGCCAGCGCTTCTCCCGCGCG 380
55 | Y P L F V G F A L L R L V A F H L G L L F V W L C Q R F S R A
55 | CTCATGGCAGCAAGAGGAGCTCCGCGCGCGCGCGCGCGCTCCGCGCGCGCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 475
86 | L H A A K R S S G A A P A P A S A S A P A P V P G G E A E R V R
86 | AGTCTTCCACAAAACAGGCTTCGAGTACATCTCCATTGCCCTGCCCATCGATGAGGATGAGAAAGTACGACAGAAGGAGCAAGCTGTGGAATGGT 570
118 | V F H K Q A F E Y I S I A L R I D E D E K A Z G O K E Q A V E W Y
118 | ATAAGAAAGGTATTGAAGAACTGAAAAAGGAATAGCTGTATAGTTACAGGACAACTGAACAGTGTGAAGAGCTAGACGCTTCAAGCTAAA 665
150 | K K G I E E L E K G I A V I V T G Q G 3 E Q C E R A R R L Q A K
150 | ATGATGACTAATTGGTTATGCCCAGGACCGCTTACAACTTCTAGAGAGATGCAACCACTTTTGCCATTTTCCAAGTCAAAAACGGACGCTA 760
181 | M M T N L V M A K D R L Q L L E K M Q P V L P F S K S Q T D V Y
181 | TAATGACGACTAACTGGCATGCCAATGGACATCTCCAGTCAAGTGGAGCTGTTCAAAAAGAAAAGCCCCCTTAACACACACTAGTA 855
213 | N D S T N L A C R N G H L Q S E S S G A V P K R K D P L T H T S N
213 | ATTCACTGCCCTGTTCAAAAACAGTTATGAAAACCTGGATCTGCAAGCGCTTTCAGGCCACCATAGAGCACTAGTTACAGTGGTTTATCCATGGTT 950
245 | S L P R S K T V M K T G S A G L S G H H R A P S Y S G L S M V
245 | TCTGGAGTGAACAGGAGTCTGGTCTCTCTCTACCCTCAATAAGTACTCCGAAAACAAATAGGACAAATAAACCTTCTACCCCTACAACCTGC 1045
276 | S G V K Q G S G P A P T T H K G 6 T P K T N R T N K P S T P T T A
276 | TACTCGTAAGAAAAAGACTTGAAGAACTTTAGGAATCTGGACAGCAACCTTGCCTAACCTTATAATGAATGAAATTTGGGACAACTGGAACAGCTG 1140
308 | T R K K K D L K N F R N V D S N L A N L I H N E I V D N G T A V
308 | TTAATTTGATGATATAGCTGGTCAAGACTTGGCAAAAACAGCAATGCAAGAAATGTTATTCTTCTCTCTGAGGCGCTGAGTTGTTACACGGG 1235
340 | K F D D I A G Q D L A K Q A L Q E I V I L P S L R P E L P F T G
340 | CITAGAGCTCTGCGCAGGGCTGTTACTCTTGGTCCACCTGGGAAAGGGAAGACAATGCTCTCTAAAGCAGTAGCTGAGCAATCGAATGCAAC 1330
371 | L R A P A R C L L L F G P P G N G K T M L A 9 K A V A A E S N A T
371 | CTCTTTAATATAAGTCTGCAAGTTTAACTTCAAAAATAGTGGGAGAAGGAGAGAAATTTGGTGAAGGCTCTTTTGTGTGGCTCGAGAACTTC 1425
403 | F F N I S A A S L T S K Y V I O G E G E K L V R A L F A V A R E L Q
403 | AACCTTCTATAATTTTATAGTTGAAGTTGATAGCTTTTGTGTGAAGAGAAGAGAAGGAGCAGATGCTAGTAGAGCGCTAAAAACCTGAATTT 1520
435 | P S I I F I D I I E V D S L L C E F R E G E H D A S R R L K T E F
435 | CTAATAGAAATTTGATGGTTACAGTCTGCTGGAGATGACAGAGTACTTGTAAATGGGTGCAACTAATAGGCCACAAGAGCTTGATGAGGCTGTTCT 1615
466 | L I E F D G V I 20 S A G D D R V L V M G A T N R P O E L D E A V L
466 | CAGTGGTTTCACTAAACGGGTATATGTGCTTTACCAAAATGAGGAGTCAAGACTACTTTTGCTTAAAAATCTGTTATGTAAACAAGGAAAGTCCAT 1710
498 | P I 3 R F I K P V Y V S L P N E E T I 4 R L L L L K N L L C K O G S P L
498 | TGACCCAAAAAGAACTAGCACAACCTTGCTAGTACATGACTGATGGATACCTAGGAAGTGACCTAACAGCTTTGGCAAAAGATGCAGCACTGGGTCT 1805
530 | T Q K E L A O L A R M I S T D G Y S G S D L T A L A K D A A L G P
530 | ATCCGAGTAACTAAACCAAGCAAGGTGAAGAAATGTCTGCTGCCAGTGAATGAGAAATATTGATTTATCTGACTTCACTGAATCTTTGAAAAAAT 1900
561 | I P E I 6 L K P E Q V K N H S A S E M I 7 R N I R L S D F T E S L K K I
561 | AAAACGCAGCGTCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGAGATACCACTGTTTAAAGAAATACCTTTGTAAACC 1995
593 | K R S V S P O T L E A Y I R W N K D F G D T T V *
593 | TGCAGAACATTTTACTTAAAAAGAGCAAAACACAAGATCTTCAATGAACGCTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTATAGAGTCT 2090
TACATATTTGTGCACCAACTTGAAGATGAACCAAGAAACAGACTTAAACAAATATACAAATGCAAAATGTAATTTTGTGTTTAAAGGCGCTTGC 2185
CTTGATGCTCAGCTTATCCCAATGGACACTAAGTTAGAGCACAACAAACCTGATTTCTGGTCTTCTTTACCAATATAATCATAATGTAATAAT 2280
AATTTGTATATTTGTGTCAGATGAAAGTATTCAGGAACAGTGAATGGTGAAGACACAAGAACATTTGTTTGTGTTCTGATGTTTTTTC 2375
TTAAAAATAGTAATTTCTCTACTTTTCTTTTCTTAACTACAGGTGATTGGAAATGCCAAACACTCTTAAGTTTATTTCTTTTTC 2470
GTTTATAAATTCAGTGTGCCAAATGAACTTTTCTCTAAGTAACTGTAATAGGAAAAAGTTTATTTGAGAGTTTCTTCTTATAAATCTACA 2565
GACATTAAACAATTTGTGTTCTTTTACCTTTTATTCTATTACTTGTCTACCAACAGTTTATAGTAGCAATATAATAGCAAAAAAGCAAA 2660
TATGGTAAAAATAGAGAAGTTTGAAGTTTGAAGTTTACTCTGTATATAACATGTAGATCAGTCTTCTATGACCTGCAGTATTTTTTTTCTAAT 2755
GTATTTGTGCAAAATCTGTGTAGACTGTTAACCTTCTCTGATGGAATTTATTTCTGCAAGAAATTTCTGATATTTAAGAGAGCCAAATTTTA 2850
ACTGCTGTGAAATGTTTCCAGTGAAGAGAAAGGAAATACTAGGAACCTAAGCAATTTCTAATTTATGCTTATTACTTTCTAATTTTACAGGA 2945
TAATTTATAAGCAAGTGAACCTACCATCTTTTATCTTAATAATTTAATCCCTCAATGAACTTTAAAAAACTGAATTTTATACATGGCAT 3040
ACATTTTCTAGTCTCTCTGCTTGTATTAATCTCAAAAGTTCTAGTTCTAGTCTGTGTATCTGCGCTTTTGTCTCCCAAAATGTACAGTAAT 3135
TCCATTTGTTGTATAAATATGCTGGATTTTATTATAAAATGTCATTGTAGGGAGTAGAGACTCATATCATGGCCTTTTAAATATGTAATA 3230
AAGGCAAAATAGATATTTGCCCTTAGTTTACTGG 3263

FIGURE 2

FIGURE 3A

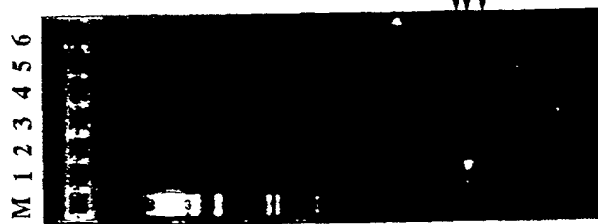
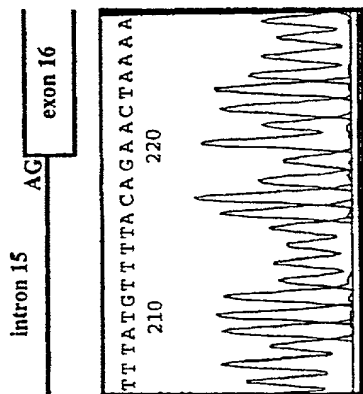


FIGURE 3B
normal splice acceptor site



mutated splice acceptor site

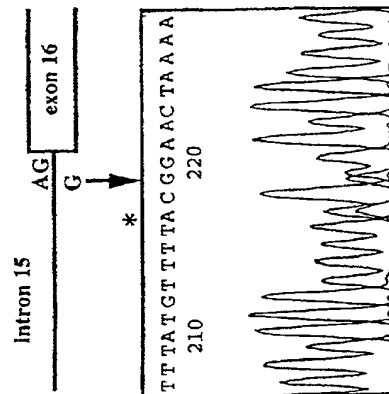


FIGURE 4A

```

AFG3_YEAST ---MNMWQYARGAPR-----SLTSL-----VCKASHISTVKP-----VLRSMPVBRLOTLSGLATE-----RTIBR--S
RCAL_YEAST -MILLSWSRIATKVVRRPFRSTYGLTRIXSLNTQTRLLNRLNENKSCNKHEDNWDADKLNKEIPTDEEVEAIRKQVEKVIQZQTKWTTIPAMWX
PARAPLEGIN MAVLLLLLRALRRGPGCPFR-----FLWGPGFAMS-PGTZARFGRGRF-----YMASRPPGD--LAEAGGRALQ-----S
YME1_YEAST ---MNVSKILVSPPTVTN-----VLRTFAPRLPQICASLLVOKKWA-----LASKKFRFYSKMSCEMPF-----K
SPAST_HUMAN ---MNSPGCGKKKGGSGGASH-----VVPFRFPFLCLATAPFAAGTAPFP-----ESPEKRNLTFTSTPLVGFALLLVLVAFBLGLLVMLC

AFG3_YEAST TQIRSFHISWTRLN-----EHRFHE-----GEGKPRGNKDMNSNKEDGKDRNEFGSL-----EYFSKEFIANTHFLTIGFTIIF
RCAL_YEAST EOKRKIDESIRLEDAVLKQESNRIOEKREKEEEMGCPXAKSNPTKEOGYFCGNNSNIFPPPPPPPPFLNDPSNPNVSKNHLFOIGLITFFL
PARAPLEGIN LQRLRLTPTTFEGINGLLKOE--LVQNFVRLWQLLGGTTFYNTSKLKQKNKEKDKSKGRAPEDDE-----EERRERDDQMYRERLRTLLV
YME1_YEAST KEADSSCKASNKST-----ISSIDNSQT--PFSRTMDRTKOANVAVSHAMLATREDEA-----NKDLTSPDAQAATFKLLQSNH
SPAST_HUMAN QRTSRALMAAKRSSGAAPAF-----ASASATAPVP--GGEAERVVVFHKOAFEYISIALREDEDEKAG-----OKEOAVEMWKKCTELEKCIATV

AFG3_YEAST TLLTFSSNNSGDDSRVLTGDFKTYLKEGLVSKITV--NKF-LVEAELVN-----TKQVVSFTIGSV---DIFEEQ---MDIOQDLNLI
RCAL_YEAST LSFLLDLNLSLEQOS-EITHODYERKLLAEGYVALLIV--NKS-MVKVMLNDONGRQADNIGRNFYFTIGSI---DSFEK---LQKAQPELDI
PARAPLEGIN TAVVMSLLHALSTSGGSISSMNDVREMLARGVQVQV--FESDVEVYLBPQAVVFCAPRLALMYRMQVANI---DKFEK---LRAADELNI
YME1_YEAST PQIVVSRFETPGIASSPECMELYHEALORIGRSEADA--QW--LLTASSAG-----AVNPSLASSSSNQ---SCYBGW---FPMYSPLYG
SPAST_HUMAN IVTGQGEDCERARRLOAKMMTNLVHAKDRLOLLEKHQT--LTFSSKQTDVYNDSTN--LACRMGLQSESGAVFRKXUPLHTMSNLPKSTVMKT

AFG3_YEAST PFRDRIPIKITYERSSPFTFLFFP-----LPTIILLCGLYFITERKINSPPNANCGCGGGLGCMFNVGKSBAKLFNKETDIK---IS--KNVA--CDE
RCAL_YEAST DXDFRIFVLYVQECNMAKAMFOJ-----LPTVLMIAIGIIMLTRS-----AQAGGSRGGIFGLSRKAKKFTETDVK---IX--KDVA--CDE
PARAPLEGIN EAKDRIPVSVYKRTGFFGNALYSV---GMTAVGLAILNTYFRLAG-----MTCREGCFSAFNOLKMARFTIVDGKMGK---VS--KDVA--CDE
YME1_YEAST SKKE--FLWVVVSESTITVYSKW---VWMLLVGILITSTFSG-----PKYITENTYLLKSSSEVADKSVVDARTN---VK--DDVC--CDE
SPAST_HUMAN GSACLSGBBNAFYSGLSMVSCVKGCGGAPFTEKGTFFKNRTNKPSTPTTATRKKKDLKNFRNVDSHLANLIMHEIVDNGTAVK--DDIA--QDL

AFG3_YEAST KOEIMTIFFLKN--GKITKLGAKIPKAI--S--T--L--K--T--C--ANVP--LSVSGSEFVEMFV--V--ARV--D--TO--SMA--S--IF--I
RCAL_YEAST KEEIMTIFFLKN--GKITKLGAKIPKAI--S--T--L--K--T--C--AGVP--YFVSGSEFVEMFV--V--ARV--D--KT--EMA--S--VF--I
PARAPLEGIN KLEVR--E--DYLS--ETILOLCAKVPKAL--L--C--L--K--V--T--AQVP--LAMAGPEFVEVIG--L--ARV--S--KE--SEA--C--VI--I
YME1_YEAST KAELEI--D--FLKD--KYESLCGLPKVL--I--T--L--K--T--C--AGVD--FFMGSSEFDEVV--V--AKRI--D--AQ--SEA--A--IF--L
SPAST_HUMAN KORLOI--LLFSLR--ELFTGLKAPA--E--LL--T--M--K--V--A--SNAT--FNISAASLTSKYV--L--EKLY--A--AV--BLO--S--IF--V

AFG3_YEAST AIGKE--GKGGALGGANDEREATLNQ--V--M--T--TS--DOVVYLAG--RFDV--N--L--N--PC--EDRBIQDS--DVNG--QOYIVLVLKRLWLDPLLT
RCAL_YEAST AIGKA--QK--WFSKANDERENTLNQ--V--M--T--T--P--DOVVYLAG--RFDI--K--L--L--PC--EDRBIQDS--ELCQ--KAIFAVLVLKRLKAGEIF
PARAPLEGIN AVGKA--ST--MSGF--SNTTZZOTLNQ--V--M--T--T--P--DOVVYLAG--RADI--C--L--M--PC--LDRBVVDL--TLQF--REFFQRLKSLKTQSST
YME1_YEAST AIGCK--NPK--DOAYATOTLNQ--V--L--F--SOT--SCIIIGCA--FFEA--K--LT--PC--YDKVVVDL--DVNG--ADILKXNKKXITLADNVD
SPAST_HUMAN SLCT--REG--EHDASRLKTEF--I--T--V--Q--SAG--DRVLVHGA--RPOE--E--VL--FIRKVVVSL--NKEET--LLLLKXLLCKQCSPLTOR

AFG3_YEAST DDMNHLSCGLT--L--P--FT--ADIANACNE--LIAARENDPTITIRBFEO--IERVIAGLEKKTIV--SKEEKRSVAYBEACH--VCGWFLKYADPLLV
RCAL_YEAST DLXN--R--L--AL--F--S--ADIANVCMNE--LIAARSDEDAVLNBFEO--IERVIGGVVERKSL--SPELKKVVAYBEACH--VCGWFLRYADPLLV
PARAPLEGIN FYSO--R--L--EL--F--S--ADIANICNE--LRAAREGHTSVETLNFEY--VEPULACTARESKI--SKEEQVVAYBEACH--VCGWFLRYADPLLV
YME1_YEAST FTIL--R--G--PL--S--ADIANVCMNE--LYACQKMAVSVDHSHFEM--KDFILMGAERKTHV--TDAAPKATAFBEACH--IMAKYTGATFLYKA
SPAST_HUMAN ELAQ--L--R--M--D--VS--SDTLALAKD--LCP--IRFLKPEOVKN--MS--SEMNRNRLSDFTES--KKIRK--SVSPOTLE--TIRWKNKDFGDTTV--

AFG3_YEAST SIIPRGQCALGYAOYLFPDQYLSIEOFRERHIMALCGRVSEELRFP--SVTSGABDDFKKVTOMANAMVTSLGMSPRICYLSFQNDG--NFKV
RCAL_YEAST SIIPRGQCALGYAOYLFGDIFLLTEOULKDRHTMSLGGRVSEELRFP--SVTSGASDDFKKVTSMATAMVTELGMSDKICWVNYKRD--DSDL
PARAPLEGIN SITPRTNAALGFAPOLPRDOHIFTEOULFERHCHALCGRASEALSPN--FVTSGAODDLKVVTRIAL--NWKDQGMAPGICGIFSFPEAQEGLMGIG
YME1_YEAST TILPRGRALGIIIFQLEMDKVDITERECQAKRLDVCMGCKIAEELLYGKDNNTSCGSDLSQATGTARAMVTOYCMSDDVCVPLVLSNM--
SPAST_HUMAN

AFG3_YEAST NKPTSNKNTARTIDLEVFSIVDDAHBACTTLLTFNLDKVDLVAKFLLRKEAITREDMIRLLGFRP--FKRPNHAFERYLD-----PKSNTTF
RCAL_YEAST TKPFSDETGDIIDSEVYELVQECNDRCTKLLKERAEDVEKIAOVLLKRFVLTREDMIDLLGKRP--FPEKADAFRYLWDETETKIRKKEEKNEK
PARAPLEGIN RRPFSQGLQOMMDHEARELLVAKAYHTEKVLQDMLDKLOALANALLEKEVINYEDIEALIGTFPGYKMIAPQRMWDAQ--KQDLCGFETTEET
YME1_YEAST -ESMSNKIKDIADNEVELLLKDSERAPKILTKKRVLLBRALGLIETTLDAHEIFQVCRCKFK--LKLFTTSTNVVVEGPD--S--DERKDIGDDK
SPAST_HUMAN

AFG3_YEAST FEAPAATN-----
RCAL_YEAST NEPKPSTN-----
PARAPLEGIN QOYPLGGFFTPPY
YME1_YEAST PKIPTMINA-----
SPAST_HUMAN -----

```

FIGURE 4B

Human: 1 [...] 459
 Mouse: 1 AGGCCGAGCGCTCCGCGCTCTCCACAAGCAGGCCCTTCGAGTACATCTCCATTGCCCTGC 60
 Human: 460 AGGCCGAGCGCTCCGCGCTCTCCACAAGCAGGCCCTTCGAGTACATCTCCATTGCCCTGC 519
 Mouse: 61 GCATCGACGAGGAAGAGAAAGCAGGACAGAAAGCAAGCTGTGGAATGGTATAAGAAAG 120
 Human: 520 GCATCGATGAGGATGAGAAAGCAGGACAGAAAGCAGGCAAGCTGTGGAATGGTATAAGAAAG 579
 Mouse: 121 GTATCGAAGAACTGGAAAAAGGAATCGCTGTATAGTTACGGGCCAAGGTGAACAGTATG 180
 Human: 580 GTATTGAAGAACTGGAAAAAGGAATAGCTGTATAGTTACAGGCAAGGTGAACAGTGTG 639
 Mouse: 181 AAAGAGCTAGACCTCTCAAGCCAAATGATGACTAATTTAGTTATGGCCAAAGGACCGCT 240
 Human: 640 AAAGAGCTAGACCGCTTCAAGCTAAATGATGACTAATTTGGTTATGGCCAAAGGACCGCT 699
 Mouse: 241 TACAACTTCTAGAGAGCTGCAACCAGTTTGGCAATTTCCAAAGTCACAGACGGAGCTCT 300
 Human: 700 TACAACTTCTAGAGAGATGCAACCAGTTTGGCAATTTCCAAAGTCACAAACGGAGCTCT 759
 Mouse: 301 ATACGAGAGTACTAACCTGACATGCCCAATGGACATCTCCAGTCAGAAAGTGGAGCAG 360
 Human: 760 ATAAAGCAGTACTAATTTGGCATGGCCAAATGGACATCTCCAGTCAGAAAGTGGAGCTG 819
 Mouse: 361 TTCGGAAGGAAAGACCGCTTAACACATGCTAGTAATTCATTGCTCGATCAAAAACGTG 420
 Human: 820 TTCGGAAGGAAAGACCGCTTAACACACACTAGTAATTCAGTCTGTTCAAAAACAG 879
 Mouse: 421 TCCTGAAAAGTGGCTCCGACGGGCTCTCCGGTCACCAAGGCGGCTAGTTGCAGTGGTT 480
 Human: 880 TTATGAAAAGTGGATCTGACGGGCTTTCAGGCCACCATAGAGCACTAGTTACAGTGGTT 939
 Mouse: 481 TGTCCATGGTTTCTGGAGCAAGACCGGCACTGGTCTGAGTACACACATAAGGGTA 540
 Human: 940 TATCCATGGTTTCTGGAGTGAACAGGGATCTGGTCTGCTCTACCACTCATAGGGTA 999
 Mouse: 541 CTCGAAAACCAATAGAACCAACAACTTCTACTCCCAACTGCAGTTCGGAAGAAAGA 600
 Human: 1000 CTCGAAAACCAATAGAACCAATAACTTCTACCCCTACAAGTCTACTCGTAAGAAA 1059
 Mouse: 601 AAGACTTGAAAAATTTAGGAATGTGGACAGCAATCTGCTAACCTTATATGAATGAAA 660
 Human: 1060 AAGACTTGAAAGATTTAGGAATGTGGACAGCAACTTGTAACTTATATGAATGAAA 1119
 Mouse: 661 TTGTTGCAATGGGACAGCTGTTAAGTTTGTGACATAGCCGGGACAGGAGCTGGCAAGC 720
 Human: 1120 TTGTTGCAATGGGACAGCTGTTAATTTGATGATATAGCTGGTCAAGACTTGGCAAAAC 1179
 Mouse: 721 AAGCGCTGAGGAGATTGTCATCTTCTCTCTCTCGGGCTGAGTTGTTACAGGGCTCA 780
 Human: 1180 AAGCATTGCAAGAAATGCTTATCTTCTCTCTCTGAGGCTGAGTTGTTACAGGGCTTA 1239
 Mouse: 781 TGCTCTCTGCTGAGGGCTGTTACTCTTGGTCCGCAAGGAAACGGAAGAAATGCTGG 840
 Human: 1240 GAGCTCTCTGAGGAGGCTGTTACTCTTGGTCCACTGGGAATGGGAAGCAATGCTGG 1299
 Mouse: 841 CTAAAGCACTAGCTGACAGATCTAATGCGACCTTTTCAACATAAGTGTGCGAGTTAA 900
 Human: 1300 CTAAAGCACTAGCTGACAGATCTGAATGCAACCTTCTTAAATATAAGTGTGCAACTTAA 1359
 Mouse: 901 CTTCAAAATATGTGGGAGAGGAGAGAAATGCTGAGAGCTCTCTTGTCTGTGGCTCGAG 960
 Human: 1360 CTTCAAAATACGTGGGAGAGGAGAGAAATGCTGAGGCTCTTTTGTCTGTGGCTCGAG 1419
 Mouse: 961 AACTTCAACCATCTATAATTTTATAGATGAAGTTGACAGTCTTTTGTGAGAGAGCGG 1020
 Human: 1420 AACTTCAACCTCTATAATTTTATAGATGAAGTTGATAGCTTTTGTGCAAGAGAG 1479
 Mouse: 1021 AAGGGGAGCAGCAGCTAGCAGACCGCTAAAGACGGAATTTTAAATAGAATTTGACGGGC 1080
 Human: 1480 AAGGGGAGCAGCAGCTAGTACAGCGCTAAAGACGGAATTTTAAATAGAATTTGATGGTG 1539
 Mouse: 1081 TGCATCTGCTGAGATGACAGAGTACTTGAATGGGTGCAACTAACAGGCCCAAGAGC 1140
 Human: 1540 TACAGTCTGCTGAGATGACAGAGTACTTGAATGGGTGCAACTAATAGGCCCAAGAGC 1599
 Mouse: 1141 TTGATGAAGCTGTCTCAGGCGTTTCAITAAACGGGTATATGTGCTTACCAATGAGG 1200
 Human: 1600 TTGATGAGGCTGTCTCAGGCGTTTCAITAAACGGGTATATGTGCTTACCAATGAGG 1659
 Mouse: 1201 AGACAAGACTCTCTGCTTAAACCTGTGTGTAACAGGAAGTCCACTGACCCAAA 1260
 Human: 1660 AGACAAGACTCTCTGCTTAAACCTGTGTATGTAACAGGAAGTCCACTGACCCAAA 1719
 Mouse: 1261 AAGAATCGCAGAGCTTGCTAGAATGACCGATGGATCTCGAAGTGATCTGACCGCTT 1320
 Human: 1720 AAGAATCGCAGAGCTTGCTAGAATGACTGATGGATCTCAGGAAGTGAACCTAACAGCTT 1779
 Mouse: 1321 TGSCCAAGGATGACCGCTGGTCTTATCCGAGAACTGAAGCCAGAGCAGGTGAAGAAATA 1380
 Human: 1780 TGGCAAAAGATGACGACTGGTCTTATCCGAGAACTAAACAGCAAGGTGAAGAAATA 1839
 Mouse: 1381 TGTCTGCCAGTGAGATGAGAAATATTCGATTATCTGACTTACAGAAATCTTAAAAAGA 1440
 Human: 1840 TGTCTGCCAGTGAGATGAGAAATATTCGATTATCTGACTTACCTGAATCTTGAAGAAA 1899
 Mouse: 1441 TAAACCGAGTGTGAGTCTCAGACCTTAGAAGCATACATAGCTGGGAACAAAGGATTTG 1500
 Human: 1900 TAAACCGAGCTGAGGCTCAGCCCTCAAACTTAGAAGCGTACATACGTTGGAACAAAGGACTTG 1959
 Mouse: 1501 GAGACCACTGTTTAAAGGAAT 1523
 Human: 1960 GAGATACCCTGTTTAAAGGAAT 1982
 Mouse: 1524 GGATGCTCTGTGAGCCCATAGAACATCGCACTTACAGGAAACAAGAGCTTTGGCTACA 1583
 1584 GGAACCCAGACTTGTGTACAGGAGCTTTAGAGTTTCAATTTTGTGCAACCAACTTGA 1643
 1644 AGAGGAACAAGAGACAGACCTAAATAAATATGCAATATGAATGC 1689

FIGURE 5

FIGURE 6B

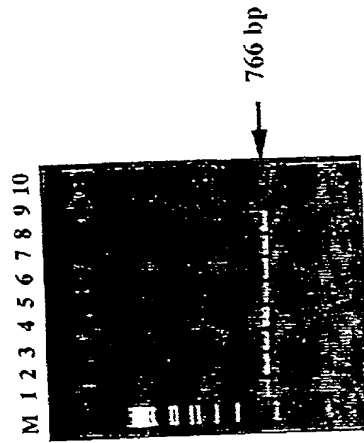


FIGURE 6C

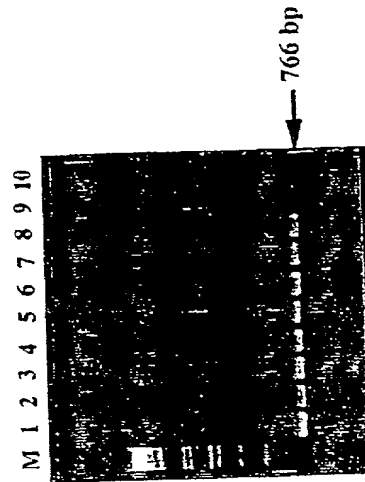
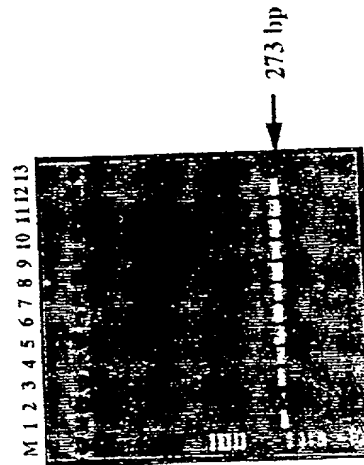


FIGURE 6A



COMBINED DECLARATION AND POWER OF ATTORNEY
FOR UTILITY PATENT APPLICATION

Attorney's Docket No.

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name;

I BELIEVE I AM THE ORIGINAL, FIRST AND SOLE INVENTOR (if only one name is listed below) OR AN ORIGINAL, FIRST AND JOINT INVENTOR (if more than one name is listed below) OF THE SUBJECT MATTER WHICH IS CLAIMED AND FOR WHICH A PATENT IS SOUGHT ON THE INVENTION ENTITLED:

CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE RESPONSIBLE FOR THE
MOST COMMON FORM OF AUTOSOMAL DOMINANT SPASTIC PARAPLEGIA

the specification of which

(check one)

☐

is attached hereto;

☒

was filed on SEPTEMBER 4, 2000 as

International Application No. PCT/FR00/02433

and was amended on _____;
(if applicable)

I HAVE REVIEWED AND UNDERSTAND THE CONTENTS OF THE ABOVE-IDENTIFIED SPECIFICATION, INCLUDING THE CLAIMS, AS AMENDED BY ANY AMENDMENT REFERRED TO ABOVE;

I ACKNOWLEDGE THE DUTY TO DISCLOSE TO THE OFFICE ALL INFORMATION KNOWN TO ME TO BE MATERIAL TO PATENTABILITY AS DEFINED IN TITLE 37, CODE OF FEDERAL REGULATIONS, Sec. 1.56 (as amended effective March 16, 1992);

I do not know and do not believe the said invention was ever known or used in the United States of America before my or our invention thereof, or patented or described in any printed publication in any country before my or our invention thereof or more than one year prior to said application; that said invention was not in public use or on sale in the United States of America more than one year prior to said application; that said invention has not been patented or made the subject of an inventor's certificate issued before the date of said application in any country foreign to the United States of America on any application filed by me or my legal representatives or assigns more than twelve months prior to said application;

I hereby claim foreign priority benefits under Title 35, United States Code Sec. 119 and/or Sec. 365 of any foreign application(s) for patent or inventor's certificate as indicated below and have also identified below any foreign application for patent or inventor's certificate on this invention having a filing date before that of the application(s) on which priority is claimed:

COMBINED DECLARATION AND POWER OF ATTORNEY			Attorney's Docket No.
COUNTRY/INTERNATIONAL	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED
FRANCE	99/11097	03 SEPTEMBER 1999	YES <u>X</u> NO <u> </u>
			YES <u> </u> NO <u> </u>

I hereby appoint the following attorneys and agent(s) to prosecute said application and to transact all business in the Patent and Trademark Office connected therewith and to file, prosecute and to transact all business in connection with international applications directed to said invention:

William L. Mathis 17,337 Peter H. Smolka 15,913 Robert S. Swecker 19,885 Platon N. Mandros 22,124 Benton S. Duffett, Jr. 22,030 Joseph R. Magnone 24,239 Norman H. Stepno 22,716 Ronald L. Grudziecki 24,970 Frederick G. Michaud, Jr. 26,003 Alan E. Kopecki 25,813 Regis E. Slutter 26,999 Samuel C. Miller, III 27,360	Ralph L. Freeland, Jr. 16,110 Robert G. Mukai 28,531 George A. Hovanec, Jr. 28,223 James A. LaBarre 28,632 E. Joseph Gess 28,510 R. Danny Huntington 27,903 Eric H. Weisblatt 30,505 James W. Peterson 26,057 Teresa Stanek Rea 30,427 Robert E. Krebs 25,885 Robert M. Schulman 31,196	William C. Rowland 30,888 T. Gene Dillahunt 25,423 Anthony W. Shaw 30,104 Patrick C. Keane 32,858 Bruce J. Boggs, Jr. 32,344 William H. Benz 25,952 Peter K. Skiff 31,917 Richard J. McGrath 29,195 Matthew L. Schneider 32,814 Michael G. Savage 32,596 Gerald F. Swiss 30,113
--	--	--

and: _____

Address all correspondence to: Norman H. Stepno
Burns, Doane, Swecker & Mathis, LLP
P.O. Box 1404
Alexandria, Virginia 22313-1404

Address all telephone calls to: Norman H. Stepno at (703) 836-6620.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor WEISSENBACH, Jean

Inventor's signature [Signature] 17 August 2001
Date

Residence 163, rue de Vaugirard 75015 PARIS / FR FRX

Citizenship French

Post Office Address the same as Residence

Full name of second joint inventor, if any HAZAN, Jamilé

Second inventor's signature [Signature] 17 August 2001
Date

Residence 22, rue des Cordelières 75013 PARIS / FR FRX

Citizenship French

Post Office Address the same as Residence

Full name of third joint inventor, if any _____

Second inventor's signature _____
Date

Residence _____

Citizenship _____

Post Office Address _____

09/830902

1

JC08 Rec'd PCT/PTO 02 MAY 2001

SEQUENCE LISTING

<110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE - CNRS

<120> CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4
GENE RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL
DOMINANT SPASTIC PARAPLEGIA

<130> D18374

<150> FR 99 11097

<151> 1999-09-03

<160> 107

<170> PatentIn Vers. 2.0

<210> 1

<211> 110000

<212> DNA

<213> Homo sapiens

<220>

<221> exon

<222> (9932)...(10471)

<220>

<221> intron

<222> (10472)...(33718)

<220>

<221> exon

<222> (33719)...(33805)

<220>

<221> intron

<222> (33806)...(35748)

<220>

<221> exon

<222> (35749)...(35832)

<220>

<221> intron

<222> (35833)...(45022)

<220>

<221> exon

<222> (45023)...(45118)

09/830902

```
<220>
<221> intron
<222> (74706)...(82788)
```

<220>
<221> exon
<222> (82789)...(82864)

<220>
<221> intron
<222> (82865)...(83102)

<220>
<221> exon
<222> (83103)...(83194)

<220>
<221> intron
<222> (83195)...(83334)

<220>
<221> exon
<222> (83335)...(83414)

<220>
<221> intron
<222> (83415)...(88129)

<220>
<221> exon
<222> (88130)...(88172)

<220>
<221> intron
<222> (88173)...(89561)

<220>
<221> exon
<222> (89562)...(89641)

<220>
<221> intron
<222> (89642)...(91162)

<220>
<221> exon
<222> (91163)...(91233)

<220>
<221> intron

106707-2952250

<222> (91234)...(93443)

<220>

<221> exon

<222> (93444)...(93484)

<220>

<221> intron

<222> (93485)...(100599)

<220>

<221> exon

<222> (100600)...(102009)

<400>	1						
taactgactc	tgctgtgttt	ccttggcatt	atagctaata	aaattgagca	ggtcaggtaa		60
cagtttatac	ttacacctac	tatttcaaaa	ccatgagctc	attcacattt	tcaactgaagt		120
aacaaatcct	ccataaacta	gaaaatctca	aactgggtgac	tgggagtttt	ggttttgttt		180
ttttgttggt	ttatttttatt	ttatttttatt	ttctagatgg	agtcttgctc	tgttgcccag		240
gctggaatgc	aatggcatga	tctcagctca	ctgcaacctc	cacctttcgg	gttcaagcaa		300
tcctcctgcc	tcaatcttcc	aagtagctgg	gactacagga	atgagctgcc	gcacctggcc		360
tggtttgttt	gtttttaaat	tttgaggcca	ggtgcagtgg	cccatatctg	tgatcccagc		420
actttgggag	accaaggcag	gccgattact	tgaggtcagg	agttcaagac	cagccaggcc		480
aacatggtaa	aaccatgtcg	ctactaaaaa	cacaaaaatt	aggccgggca	cggtggctca		540
cgtctataat	cccagcactt	tgggaggcca	aggtgagtgg	atcacctgag	gtcaggagggt		600
caagaccagc	ctggcaaaca	tggtgaaacc	ccgtctctat	taaaaataca	aaaattagcc		660
gggtgtggtg	gcgggcacct	gtaatcccag	ctattgggga	ggctgaggca	ggagaatcgc		720
ttgaacctgg	gaggtggagg	ttgcagtgag	tcgagatggt	gccactgacc	tccagcctgg		780
gcaataagag	cgagacgcca	tctcaaaaat	atataaataa	ataaataaaa	atacaaaaag		840
tagctaggct	ttgtggcagg	cacctgtagt	cccagctact	taggaggctg	aggcaggaga		900
attgcttgaa	cccaggaggc	ggaggttgca	gtgagccaag	attgtgccac	tgcgttccag		960
cctgggtgac	agagcaagac	tccttctcaa	aaataaataa	ataaataaat	aaataaataa		1020
ataaataaaa	ttttgagctg	ggcatgaaag	ctgaggcagg	aggatccctt	gagcccagca		1080
gtttgagacc	ccagtgagct	ataattctga	cactgcactt	cagcctggct	gacagaggga		1140

gaccgtgtat	ctaaaaagaa	taaaaataaca	atgattttttg	agccaataaac	tcttagccaa	1200
tagatttcac	ataaaaaattt	agaatttctgg	tttctcttga	aaaattaaaa	aaatctgaca	1260
atgttaagct	tcacattcct	gaaaggcaaa	aatcagtgga	agctagctgg	gtgctgtggc	1320
tcacgcctat	aaacccagtg	ctttgggagg	ctagggttaag	tggattggtt	gagtccagga	1380
gtttgagatc	agcctggcca	acacagtcaa	accccatctc	tacaaaacat	acaaaaatta	1440
gctgggatat	ggtggtgtgc	acttgtgtgc	ccaactactc	aggaggctga	ggcaggagga	1500
ccacttgagc	cctggagggt	gagggggcag	tgagccatga	ttgtgccact	gcactccagc	1560
ctggacgaca	gagagacatt	gtctcaaaaa	gaaaaaaaaa	aatcagctca	gtgggagctg	1620
agtaacagct	gtcctgttca	attacaggat	gcaactcttt	agcttctcat	agtttccatc	1680
attcacatca	tacatgcatt	tttgcattgc	acacaaccca	catatggaac	ccatatatgt	1740
tatatgcctg	accactgttg	ctattggaag	ttttggccac	tgcattaaac	tataaactcc	1800
atcttattaa	tcttgacacc	ccacttattg	cctgatatat	tgtccgtttc	ttaatattcta	1860
ttcatagaac	aatgaatga	ataatatgtg	ccacattgtg	gactcaattc	agggagatga	1920
ctaattcattc	acaattatgt	ttttccttct	taatacagag	attggataat	aattccccaa	1980
ttgttacttc	tctcatccct	cccctccaac	catatctttt	catttgtttt	tcttatttta	2040
ggttttctgc	ctttttcaaa	atcagccatt	tctcactgg	actctacatg	tgccattttt	2100
tttttttttt	ttttactaat	ttttttagtt	gaaaagaggt	ccttaatatc	tgtcattgggt	2160
ccacacttga	aatctagaat	ctctcgaatt	gaagggtctga	agagttcctt	agaaggcaag	2220
tagggccgaa	atatccaaaa	aaaactattt	gatggtaggc	actgtggtaa	ataaatatgg	2280
gttaaagtga	taaagtgacc	gaatgtcctg	gattagttag	cacagtacct	agctccttct	2340
ttttgtcctt	gtgaaactgt	ctcagattcc	attcaagatt	aagtgtcctg	aaagttctta	2400
caagctgaaa	actgggaggc	tctcacactg	taggtagaat	ggctagcagg	gttgggatct	2460
catctaggca	cattgcagaa	ccagggtaac	accaaggcat	tatttttttt	ctttctttac	2520
ttttcttctt	tttttttaga	gacagggctc	cattatgtta	cccaggctgg	attccaattc	2580
ctgagctcaa	gtgacctctc	tgagtagctg	ggacttcagg	tgtgcaccgc	tgtgtctgaa	2640
ggcactgttt	tttttgtttt	ttttgttttt	tttgacacaa	atthaggaag	atgttaattc	2700
acaacagtct	tgagactgag	atataattcc	aaggagcaga	agatgtgagt	ttagagcaact	2760
aaaaaatgag	attccagtag	agtcagaaat	ctgaaatggc	attacagata	taagagacaa	2820
aaacaaaatg	tattgagctc	tgtcatgttg	caggcatcat	gatggagggt	ttagatgtac	2880

tctttcat	ttt	tgtaattttt	atagaggaat	taactagaat	agcaacccca	gtcccactaa	2940
ctctaaatcc	catat	ttttta	ccatacaaaa	aaagagcaaa	agtgcagaaa	agcacagtca	3000
atattaagta	caaccagata	gcagagactc	agtaa	atggg	aggccggagg	cccga	3060
aacataatgc	ccatgaatga	aagcccatca	cttg	cgcttc	aggggctaac	aatatactta	3120
tttcataatt	aaaatagaaa	cagactatgt	aaaaatatta	ttcttgagat	cccagatttt		3180
taccttaaat	tactgaggca	gtaagtgtaa	ttaactaata	tgtgatgttg	ggcaaataac		3240
agactcttag	agccccaat	tctttat	ttt	aaaaactga	ggccagatga	ggggtgcttac	3300
acatgtaatc	ccagcactgg	ggatgccaaa	gcaggcagat	cacttgagct	tgagcttagg		3360
agtttgagat	cagcctgggc	aacacggtta	agacctcatc	tctaaacaaa	atacaaaaat		3420
tagccaaaca	tggtggtgtg	cacctgtgtt	cctggctact	caggaggctg	aggtggaagg		3480
atcgcttgag	cccagagagt	tgaggggtgca	gtgagccatg	atcttgccat	tgcaactctac		3540
actctagtct	gggtgacaga	gtgagactct	gtcttggggg	aaaacaaaag	agatgataat		3600
gcttaactga	agtagcaata	ttttaaaaag	gcactaaaag	ttcatctgct	tagttcagaa		3660
tatgggttcg	ataaatatta	gcaagtagta	gtagtcatca	tcatcactgt	cactgctggt		3720
ctctccttaa	acttaagcat	gttttttgtt	tttttgagac	agtgtctcac	tctgtcacc		3780
aagctggagt	tcagtgggtgt	gatcttggct	cactccaacc	tctgcctccc	aggttcaagt		3840
gattctcctg	tctcaccctc	ccaagtagct	gggaccacag	acacgtacca	caaccacacc		3900
cagctaattt	ttcgtat	ttt	tggtagaggt	ggggtttcac	catgttggcc	aggctggtct	3960
tgaactcctg	acctcaagtg	atccacctgt	ctcggcctcc	caaagtgctg	ggattacagg		4020
cgtagccac	agcatccagc	cttaagcatg	ttaattaagt	ttttataatt	cagcaaaatg		4080
gttgaaaaat	gctgtcttaa	atgagatgct	taagctgccg	tctgaacatg	aggtagaagg		4140
aaattctaca	cataatcatt	gtgctaaatt	acttgcaaag	atggccacaa	caattcctcc		4200
tatcctcata	tatatgcccc	tttgcaatgt	gactttgcta	cttctctatc	aagatgtgga		4260
gcttat	ttt	ccatatattg	cactagagtt	ggccttctga	cttgctttga	caatggaatg	4320
tagtaciaa	at	gacactgtgc	aactttggat	tttaggtttc	gagagaactt	acaccttcca	4380
ctcacactct	cttgga	aacc	agatgcaatg	taaagaagtc	agggtatcc	tgctagagac	4440
atatgtccca	gctaata	gccc	acaatcaacc	tctgaacata	tgaatgaggc	tagctaggcc	4500
atacagccat	tcgggtcaagc	catcagatga	ctacatccac	aggaatgatc	cacaggcaag		4560

gccatcagaa	gaaccatcca	gctgaactta	ccccaaattg	ctgagtcaca	aagttgtgtg	4620
taaataaatg	tctgctatct	taagccagtg	agtttttgag	tggtatatta	catagcatca	4680
gaaatctaac	acaatcatta	tgtttgaatc	atttttcaaa	tttctcatat	ttattaaatg	4740
agtaccataa	gcaaggtgtc	aggctggatg	caaaaagtga	ggcaaatgt	ataaagtgtg	4800
accactgcct	tcagtaagtt	tacaatctat	atcaagaggt	gatgaagtgt	ttaaataatc	4860
atcctgcagg	gcaatatagt	ataagagcca	cagagtaaca	caaccatatt	gtcataacaa	4920
ctgaaaaaca	agatcatttc	tgctggaggt	gataatggaa	taatttatca	agaatataac	4980
agagctggac	gcggtggctc	acacctgtaa	tcccagcact	ttgggaggcc	aaggaaggtg	5040
gatcacaagg	tcaggagttc	gagaccatcc	tggttaacac	gatgaacccg	tctctactaa	5100
aaatacaaga	aattagccgg	gcgtggtggc	acgcgtctgt	agtcgcagct	actcaggagg	5160
ctgaggcagg	agaaccactt	gaacatggga	agcagaggtt	gcattgagct	gagatcgtgc	5220
catggcactc	cagcctgggt	gacagagtga	gactcagtct	caaaaaaaaa	aaaaaaaaaa	5280
aaaatataac	attagaggta	agtcttgaag	gactttgaca	gtggaagtag	gaggcgaggc	5340
cattctaagt	gaatgaaaaa	tgacaggaga	gtaattgtag	tcctggaaaa	gagcaaagta	5400
ggtacagacc	aacagtctat	attagctaga	gtatagtga	agtgcagagg	aaatgtcgga	5460
gaaccattct	ttattcaaaa	actatcttcc	tcatggccaa	gcatagtggc	tcatgcctgt	5520
aatcccagca	ttttgggagg	tcaaggtgag	tgatcactt	gagctcagga	attcaagacc	5580
atctggggca	acatagtgag	acctcatctc	aactaaaaaa	caaaaaattc	agacagatgc	5640
agtggctcac	acctgtaatc	ccagaacttt	ggtaggctga	ggcgggcgga	tcacgaggtc	5700
aggagatcaa	gaccctcctg	gacaacatgg	agaaacccca	tctctattaa	aaatacaaaa	5760
ttagctgggc	atggtggcac	atccctgtaa	tcccagctac	tcgggaggct	gaggcaggag	5820
aatcgcttga	accagggagt	cggaggttgc	agtgagccga	gatcgcacca	ctgcactcca	5880
gtctggcgac	agagcgagac	tccatcttaa	aaaataaata	aattttaaaa	aaaactaccc	5940
cagcatggtg	gtgcatgcct	gtagtccag	ttactcagga	ggctgaggca	agagggtggt	6000
ttgagccagg	gaggtcaagg	ctgcagtgag	ctctgatggc	gccactgtac	tccagcttgg	6060
gtgacagagt	gagaccttgt	ctcaaaaaca	aaaacaaaaa	caaaaaacca	acaaatctcc	6120
ttgttagtat	catggtgagt	aaaaaataaa	ataaaaatag	aaataaactg	aacatggtgg	6180
ctcatgcctg	taatcctagc	acttttgaag	gctgaagtgg	gaggattgct	tgagggtggt	6240
agttcaaaac	tggcttgggc	aacacggtga	gagagacctt	gtctctacaa	aagaactttt	6300

tgtagggcag	tactggcggg	gaataattaa	gtaaaattgg	ctggtgctgt	ggctcatgcc	8040
tgtaatccca	gcactttggg	aggccgaggc	gggcaggttg	cttgagccca	ggaattcaag	8100
accaacctgg	gaaacatagc	aagaccctgt	ctcaacaaaa	aagtaaaaaa	attagctggg	8160
ggcgcgatgg	ggtggctcat	gcctgtaatc	ccaacacttt	ggaaggctga	ggcaggcgga	8220
ttgcttgagc	ccaggagttg	gagaccagcc	tgggcaacat	ggtgaaaccc	tggctctata	8280
aagaatacaa	aaattagtcg	ggcccagtgg	cgtgtgcctg	tgatcccagc	tactcgggag	8340
gctgaggtgg	aaggatcacc	tgagccaggg	aggtggaggt	tgcagtgagt	catgttgttt	8400
gcgccactgc	actccagcct	gggcaatgga	gtgaaaccct	gtccaaaaaa	taaaaaata	8460
aagctgtggc	agaatgtgga	gattcttgga	agctggaagc	tctcatgggg	catttgga	8520
cctcacattg	taaataacgg	agtcttttta	tcagtttggc	ttccttagtt	ttaggaaaca	8580
agaaataatt	atggctaact	caagtaaaaa	gagaaagaga	agagaaaaaa	gacgtggaga	8640
tagagagaga	gggagagaga	ggaaaagacg	aaaggaagga	agggggaaag	gagagaggaa	8700
gagagaaaca	gagaaacaga	ctgattagtg	tattggatag	attacataac	caagtgacca	8760
gtcaggaacc	cagcagctct	gggggagctc	aatgtgatgc	attgataaac	ccgctcttaa	8820
gagcactcgt	ttccagttac	tttctattcg	gtgggtctcc	agccaagatt	ccagggtcca	8880
ggagaatctg	actgacctag	tgtttgcttc	cgcttttgcg	gtctgggttc	tgtgcttgca	8940
gctcattaga	atacagggag	cagagacaag	caggtagttt	cccaaaggaa	gggatgctga	9000
gtagattaaa	aaaaaagtgt	agattcttca	gtaaactatg	ggatggtaac	tatgcaaac	9060
ctaagatttc	ccttattcaa	ataaattatc	tttcatatta	gacatctaaa	tatgcactaa	9120
tttagttaaa	cccctgggtt	agttgatctc	atcacactga	gctaacattt	ttgttgctgt	9180
tgtttgagc	gacctgaagt	ttcttatctt	cacaattgct	ttcctctcaa	ataattccca	9240
gattttaaat	ttttatttta	ttttttctgg	agacggagtc	tcgctctgtc	gcccaggctg	9300
gagtgcagtg	gcgcgatctc	agctcacttg	cagcctctgc	ctcccagagtt	caagcgattc	9360
tccggcctta	gccttccaac	cagctgggac	tacaggcgcg	cgccccacg	cccggcta	9420
ttaattccca	gattgatatc	cattgcttct	gagatgggcc	aattatcctt	cggagaagac	9480
ttaggtcgcc	tggcagaaaa	agatgaaaga	aatctaagaa	aacgacgaca	ctgagagagg	9540
agcctagcga	accagcagag	cgaccccaag	ccgcaattcc	cccttccgtg	gatcgattac	9600
gaaggcttcc	tggcaggagc	tctccagggc	tgccgacgtg	agccgaactg	cacattggga	9660
actgtagttg	agtgggaaag	ccgagaggcg	ggggccgcac	acgcgtacag	gggccccggt	9720

cgcttggtta	atTTTTgtat	tttttagtaga	gtagggattt	caccatgttg	gccaggatgg	11460
tctccatctc	ctgacctcgt	gatcgggctg	cctcggcctc	ccaaagtgct	ggtaatacag	11520
gcatgagcca	ccgcgcctgg	ccttaagcta	acatttttta	ttatatgtgc	caggcattgt	11580
gctatttagct	ttgcatgtgt	tatctttttt	ctttttaaaa	aaaatagcaa	ccatcccaga	11640
ccatgaaaag	tgttatttta	tcctcacaat	aactttgtga	gatgaaggta	ttattggtat	11700
cagcatttta	gagatgaaga	aaatgaggcc	caaaagataa	aggaggttat	tccaaacctg	11760
tgctaataat	gaaatttctt	ttatggaaaa	taagtgaat	taggagaagt	ctaactttta	11820
caattctccc	cttttatact	tacttccagt	atgctgagat	cttgcttctc	cctctgccaa	11880
aaacacccac	ttttctacca	cacctcaatt	agataactac	ttgcattgtc	cattagtga	11940
aacagaaaca	atctgcactt	cattcatagt	gtctgtctct	actgccaaact	ccaaaaactg	12000
ttctagaatt	tcaatcttgt	tgaaacctgt	ttcctttgtg	gggcctggga	gtgggaagtg	12060
ggatataagg	aaagaggcag	ttattcattg	ttttggacag	taaggaaaga	gtgacggtta	12120
agagaggcca	aggagggtgt	taacatttta	gaatactatg	tgttttaga	aggaaatttt	12180
ttgttaactg	ctccattctt	tttttttttt	ttttttgaga	tggagtctcc	ctgtgtcacc	12240
cagactggag	tgcagtggcg	tgatctcagc	ttactgcaac	ctctgcctcc	cagacgcagg	12300
cgattctctt	gcctcagtct	cccagatagc	tggaaaccaca	ggtgcgcctt	atcacggctg	12360
gctaattttt	tgtattttta	gtagagatgg	tgtttcacca	tggtggccag	actggtcacg	12420
aactcctgac	ctcaggtgat	tcgcctgcct	cgaccttcca	agtgccttga	ttacaggtgt	12480
cagccactgt	gcccacccag	caagctccat	tctttatcac	ctcttaagaa	catccaggat	12540
cccttgggga	gaattaaaac	ggttgcaaa	ttttagaata	gaggaacatg	tttaagcgta	12600
gatcattttt	tgggocagtc	atggtggctc	acgcgtgtaa	tcccaacact	ttgggaggcc	12660
ggggcggggtg	gatcatgagg	tcaagagttt	gagaccatcc	tggccaacat	ggtgaaatcc	12720
catctctact	aaaaatacaa	aaattagctg	ggcatggtga	cacatgcctg	tagtcccagc	12780
tactcaggag	gctgaggcaa	gagaagtgtc	tgaacctgga	aggtggaggt	tgcagtgagc	12840
cgagatcgtg	ccactacact	ccagcctggg	cggcagagca	agactccgtc	ttgaaaaaaa	12900
aaagtaggtc	atTTTTggct	gggcacgggtg	gctcatgcct	gtaattccag	cactttggga	12960
ggctgaggtg	ggtggattgc	ttgagcccag	gagtttgaga	ccagcctggg	caacatagtg	13020
aaaccctgtc	tttgtgaaaa	atacaaagat	tagctaggcg	cagtggcaaa	tgctgttagt	13080
cccagctact	tggggggctg	aggtaggagg	atcacttgag	ctcaggttgt	tcaggctgca	13140

atgagctgag	atcgtgcaac	tgactccga	actgggtgac	aggagtaaaa	ctgtctcaaa	13200
aaaaaaaaaa	aaaaaaacca	aaaaaaaaaa	aactgtttta	attgttttat	ttaggaagag	13260
aaagtcagaa	catgcaagga	aatttttttt	atttgtttat	ttttgagacg	gagtctcgct	13320
cagttgccca	ggctggagtg	caatggtatg	atcttggtc	actgcaacct	ctgcctcccg	13380
gattcaagcg	attctcctgc	ctcagcctcc	tgagcagctg	ggattacagg	tgtatgccac	13440
cacgcccagc	taatttttgt	gttttttagta	gagatggggt	tccaccatgt	tggccaggct	13500
ggtttcgagc	tcttgacctc	aagtgaaccg	ccctccttgg	cctcccaaag	tgtctgggatt	13560
acaggtgtga	gccgcggtga	ccgaccacaa	ggaaatttta	gttaacactg	ttggttgatg	13620
ggagttggga	ggtaggataa	aaggagaaat	taaggaaaac	ctaggcatga	aaaataaaag	13680
accctgagct	cttagatttg	aagaaatagc	agttccatgt	gaggaataag	tggaagaaat	13740
agaattcaga	cctcaggtct	caggctggtg	acttaaattc	ttcagtatca	catatatgaa	13800
tatatctatt	atacttaaat	cctcctaaac	atttttat	ttcagttgga	tatattaaat	13860
atataaaaaat	aattatttta	tttatttttag	agacagtgtt	tcactctctt	acccaagctg	13920
gagtggaatg	tgatcatagc	tcactgcagc	ctcaaggctc	attcctaggc	ttaagtgatc	13980
ctcttttttt	tttttttgtc	agagatagga	tcttgttatg	ttagccaggc	tggagaaaatt	14040
tcaacaatat	tttgaacaat	aaaaaaaaaa	taaattaggt	tttattgtaa	agtggtatgt	14100
ttaagtttac	gccattctcc	tgcgtaatgc	atatttcata	ctcttcctac	tgataatgtt	14160
tctgttcaca	actttttctt	tatatattga	tttcttttct	cttttctttt	tttttttttt	14220
tgagacagag	cctcgctttg	tcgcccaggc	tagagtgcag	tggcgcgatc	tcggctcact	14280
gctagctccg	cctcccggtt	tcacgccatt	ctcctgcctc	agcttcccga	gtagctggga	14340
ctacaggcgc	ccgccaccat	gccagctaa	ttttttgtat	ttttagtaga	gacgggggtt	14400
cactgtgtta	gccaggatgg	tctcgatctc	ctgacctcgt	gatccaccgc	cctcggcctc	14460
ccaaagtgtc	gggattacag	gcgtgagcca	ccgcgcccgc	cacatatatt	gattttctaat	14520
gtggacatca	gaatgggctg	ttgacagctc	tttttttttt	ttttgaggca	gagtctcgct	14580
tcttcaccca	ggctggagtg	caatggcgcg	atctcagctc	actgcaacct	ccacctccca	14640
ggttcaagcg	attctcctgc	ctcagcctcc	ccagtaactg	ggattacagg	catgtgccac	14700
cccacaccca	gccaattttt	gtatttttag	tagagacagc	atttcaccat	gttggccagg	14760
ctgtttctga	actcctggcc	tcaagtgatc	cactcgctt	ggcctcacia	actgctagga	14820

ttacagggtgt	gagccaccggt	gccagcctt	gacaggctctt	tagtttgatt	ttagttcaac	14880
aactgatgcc	gtaatatgcc	aaattaaatt	agttcagact	gaaacggatt	acttaaagat	14940
tcattttcct	ttaaaaatga	agtaaaaactt	tagccggatg	tggtagcgagg	tgtgtgtaat	15000
cccagctact	cgggaggggtg	aggcaggaga	atcgcttgaa	cccaggaggt	ggaggttgca	15060
gtgagctgag	atctcaccca	ttgcactcct	gcctgggtga	gaagagtgag	actccatcaa	15120
aaaaaaaaaa	aaaggctggg	catggtggct	caagccgggt	gcagtggccc	atgcctgtta	15180
tcccagcact	ttgggaggca	gaggcaggcg	gatcacttgt	agtcagaagt	tgcgaccag	15240
cctggccaac	atattagtga	aaccccgctt	ctactaaaaa	tacaaaaatt	agctgggcat	15300
agtggtgcac	gcctgtagtc	ccgggtactt	gggacgctga	cgcaggacaa	ttgcttgaac	15360
tccagagggtg	gagattgctg	tgggccaaga	tcacgccact	gcacgccagc	ctgggcaaca	15420
gagcaagact	ccatctctca	aaaaaaaaaa	aaaaagaaaa	ctgagtttat	attgttatgg	15480
ttcttatcta	tcttaaattt	tttctttagg	agattgaata	ttttgtact	ttagtcttaa	15540
tccagtgggt	gaaagagagt	gcacttttga	agtctgcctc	ttggctgtcc	ttgacaacac	15600
aaaccttagt	tccaagagaa	tgtaattcct	cctctttctc	agtgttcaa	aatatatatg	15660
atcaaataga	aacgagttga	ataggcagtc	tcttcaaagg	tttcctaact	ctgtgggttaa	15720
ctatcaagga	gctggttaata	tcatgcactg	ccattccctt	ggcaacatga	cttatctttt	15780
tttttttttt	ttcttttttg	agacagagtc	ttgctctgtc	accaggctg	gagtgcagtg	15840
gcgagatctc	agctcgctgc	aagccccgcc	tcccacgttc	acgccatttt	cctgcctcag	15900
cctcccaagt	agctgggact	acaggcgcct	gccatcacac	ctggctaatt	ttttttgtat	15960
tttttagtag	agacgggggt	tcaccgtgtt	agccagtacg	gtctcgatct	cctgtcctcg	16020
tgatccgccc	accttggcct	cccaaagtgt	tgggattaca	ggtgtgagcc	actgcgcctg	16080
gcctacttat	cttctaattt	aactgaaaac	caatttattt	gattcagtga	aatggcatca	16140
aactgtagta	gtgttaattg	aaatatttgg	tacottgaaa	tgttaaatgc	caaattaaat	16200
ctcattttta	tgaaatctcg	tgtaaatgtg	ttttatatgg	tgactatgtt	tattctgaat	16260
tttattctta	tggcatacta	aaaaaaaaaa	aaattttttt	ttttttttgt	aatggagtct	16320
tgctctgtca	cccacgctgg	agtgcagtgg	cgcaatctcg	gttcactgta	acctccacct	16380
ccggggttca	agcgattctc	ctgcctcagt	ctcccagagta	gctgggtacta	caggcggtgca	16440
ccactatgcc	tggctaatat	ttttgtattt	ttaggagaga	cagggtttca	ccatgttggt	16500
caggctgata	tcgaactcct	gacctgaagt	gatccgcctg	cctcagcctc	ccaaagtgtt	16560

gggattacag	gtatgacca	ctgcacccaa	ccataactca	aatttgacac	tgaattttca	16620
taaaggcctt	aatttatctg	aaaccaaact	atttcaaaag	aggaatagca	cagcaaattc	16680
tgttgactta	atgagaggat	atgtgaagtc	tatttattaa	agcaaattt	aattggaggc	16740
cagttaattt	gtacagctct	gcattttaga	tatttgagaa	atattttatt	cctctccagt	16800
gagatgtgtt	aaaacattag	ttatgtgatt	aacaaatatg	tgtacatacg	tatatatgta	16860
catacacatt	ttgagacagg	gccttgggtc	gttgttcagg	ctggagtgca	atgacaccat	16920
cttagctcat	tgaagcttca	gcctcgagc	ctcaatcgat	ccaccacct	cagcctccct	16980
agtagctggg	tctacagaca	tataccacca	tgcttggctg	attttttaat	tttttgtaga	17040
gatggtgatc	ttgccctgtt	gactaagttg	gtaaatattt	taattgttga	actttcttgg	17100
aagactgaaa	acctgtgata	gccattttat	ataaggagaa	gctgaagttc	aaagagtaga	17160
ctcatagcag	aaacaaaaat	agaatttaag	tgaatggact	caaaatattg	tactttttac	17220
tttatactgc	aggtttttat	gttgtaatgc	tggtaatgag	ctccttgga	tatttggagg	17280
aaaagagaag	gttgtaataa	tggttctttg	gatttactag	aacatatcat	gttctgcatg	17340
gctcctgtgg	gtagacaagc	cggaaatctc	ctgggtaaca	caatggtgga	ggttctctag	17400
gtgacgtttg	atttctcaag	tacataggac	taaacagaaa	aggcctagta	tgttatatga	17460
atgagagatc	aagtttctca	ggatattcta	gggctaaagg	atcaggcatc	gaagacagaa	17520
attgtctaaa	taaaattttt	tctattcata	gttttaaagg	gctaaaggg	caggcattga	17580
agacagaaat	tgtctaagta	aaatattttt	ctgttcatag	ttttaagctg	tgtatatgtg	17640
catgtgtgta	tttaaaataa	cttcgtcaat	gaaaagagtc	aaactctgta	aaatatttga	17700
agagatttat	tctgagccaa	atatgagtga	ccagtggccc	atgacacagc	cccagtagat	17760
actaagaaca	tctgtccaag	gtggtcaggc	tatagcttga	ttttatacac	tttagggaga	17820
cataagacgt	cagttaaaca	tgtaagatgt	acattggttc	catctggaaa	ggcaggaaaa	17880
ctagaagttg	gggaggcttc	caggctcgtg	gcagattcaa	agattttctg	attggcaatt	17940
ggtcaaaaaga	gcttatctaa	agtccctggaa	tccatagaag	ggagtgtctg	gtttaaaata	18000
ataggttgta	gctaccaagg	tttttattgt	acagatgaag	cctccaggta	gcaggcttca	18060
gagagaatac	attgtaaattg	tttcttatga	gactttaaaa	ggtggcagac	tcttaagtta	18120
attttctcct	ggttcaggta	aaagacttgg	aaagggaaag	gattctctac	agaacgtaaa	18180
ttttccccc	aagagaaagc	tttgcagggc	catttcagaa	tatgtcaaa	aaatataaatt	18240

tagggtaaaa	tacttcaatt	tgttttat	atattttat	tttttgagac	agagtctcgc	18300
tctgttgccc	aggctggagt	gcagtggcat	gatctcggct	cactgcaagc	tccgcctcct	18360
gtgttcacgc	cgttctcctg	cctcagcctc	ctgagtagct	gggactacag	gcgctggcca	18420
ccatgcccag	ctatTTTTTT	tgtatTTTTa	gtagagacgg	gatttcacca	tgttggccag	18480
gatggtctcg	atctcttgac	ctcgtgattc	acccgcctcg	gcctcccaaa	gtgctgggat	18540
tacaggcgtg	agccactgtg	cccggcctca	aaatacttca	atctctttca	tggcctgcta	18600
tctgacgtga	tgctgtacta	gagtcagget	gggaatttgg	cgtcttattg	ctacaaaaca	18660
tcttaatatc	tctgttttaa	tgtaaatgct	gatcagttgt	ccctgaattc	caaagggag	18720
agggtatatg	agtcattgtc	aacccccact	tctcattatg	gcctgaacta	gttttttagg	18780
ttacttttgg	aatgcctttg	gcaaggggag	ggccatgag	tcagttgggg	gtcttagagt	18840
ttatTTTTTT	gcttaccggt	ttataaaaaag	ttaatgaaaa	ttatcatatt	tcataattct	18900
atataattca	atattgtact	ttatttaaaa	ctcacgtata	aaatagctgt	ccatatctgt	18960
tttcagaaga	tgaggatgga	gggtagaaat	cagaagtgtc	agatttggtg	atcttcttac	19020
actgctgaaa	acctatacct	accactttga	aaggattaat	ttcagacttg	ctttcttttg	19080
gcctaattgat	tctactttga	agtttctctg	attaaactaa	ggaataaatc	tgataaatgg	19140
acattcagat	gataccatac	ttttccaaaa	gataacattg	cttttgatta	catatgcaat	19200
aaacatttca	cattttttct	caagattatt	tactggcatc	tgcaccaaag	acacaaaaaa	19260
gcagccactg	ttaaggactt	tatcccttgt	tctgttttta	gctggtttgt	tgttgtat	19320
ttcctcatgt	tgaatacagt	taaaccctat	taaactggat	tcccatatt	actgttagtt	19380
gtcctgatga	caattaagga	tagttagtaa	atggatattg	aatcatttta	tttttttagt	19440
agtgatggcg	tttcgctgtg	ttggccaggt	tggtctcgaa	ctcctggcct	caagtgatct	19500
gcctgtgttg	gcccccaaaa	gtgttaggat	tacaggcgtg	agccactgca	cctggctgga	19560
tattgaatta	ttgaagaagg	atgttcctta	acacttctcg	cctcttttcc	gctgtcttac	19620
tctctctgtt	ccaccagta	gtagttgggt	ttttgatgtc	tggcaaagta	tagattgtct	19680
actcttctat	attagttttc	ttcactcttt	tttttaaaaa	tgtttttaaa	gcaagataga	19740
gacagggtct	tgccgtattg	cccaggctgg	tcttgaactc	ctgagctcaa	gcaatcctcc	19800
cacctacccc	cataaagtgc	ggagattaca	ggcatgagcc	accatgcctg	gccagttttc	19860
ttcactcttg	acctataata	gtcctgcaaa	gccagtgaag	ctgttaatat	gctgacgtag	19920
cctttttttc	tctcgtttgt	gaattattaa	caatcgctga	tcttacacat	catatacaat	19980

aaaacatctt	tgaacactgt	cacatctcag	cagctcattc	tggttaatga	ggaaagaaaa	20040
atgtcaaaat	ctgtgatttt	cttagagggt	attaaatggt	ttacagctat	gtagatatctc	20100
tgtagacttg	ctgtactttac	ttacataact	tttctgctct	tctgcagagg	gagagaatta	20160
acttcataag	tgggctttgt	caatgccctc	cagtcagcga	tctccaggaa	caaacttatc	20220
tatgggtgag	caagtggggg	ttattaccca	ctgtagccag	ggagaacaca	catatgtaat	20280
aaccatggta	tgtcatagta	aagggtgtta	ggaagatagg	atttgggctc	gtttattttgg	20340
tgattttaag	gagggtttta	agaagcaggg	ttttgctctg	atttggatgc	taccaggagc	20400
tgcgaataat	tctatgaata	aatatcataa	aacctatcta	gaagaaaaga	ctagagtgag	20460
gtctaaagct	gtagtagtta	aaaagcaaag	gtcactcctt	atctggaaaa	ggggaatatt	20520
tggatattca	tggtttagac	agtgttcagt	gttcatgttt	tgctgtgttt	tagacataat	20580
tgtagagtgg	tcttgttcca	ccgtgggtcac	agagtggcat	tatttcatgc	tgatattctg	20640
taagtgccag	gtcaaatacc	aagggttagc	tgatagtagt	aggccagctc	ctggatgtaa	20700
taagctgttt	ttctctttct	cagctttttt	ttttgttggt	ttttttttta	aatagagtct	20760
ggctctgtca	cccaggctgg	agttcagtg	gggcaatctc	agctcactgc	agcctctgcc	20820
tcctggattc	aagcgactct	cctgcctcag	cctcccaact	agctggaatt	acaggcgcaa	20880
gtcaccacac	ccggctaatt	tttgtatttt	tagtagagac	ggggtttcac	catgttgagc	20940
aggctggtct	cgaactccgg	acctcaagca	atccatctgc	ctcagcctcc	caaagtgctg	21000
agattacagg	catgagtcac	tgcacctggc	ctgtttctta	gatttgaggg	tcaactttta	21060
cccttttcga	actgtggggc	ttcttatgga	aattgacatt	taagtctga	ccatatagga	21120
tcttgggcaa	gttacttaac	tatgtatgct	gagatgtttt	tttaaagtct	tagtgcttgc	21180
acataatagg	tactcagtaa	atgagaacta	ttattataaa	atcaatagta	cttttaagat	21240
tacagctaga	ggttatgtta	gaacattggt	aactcttcgt	actagtttct	gttacacttt	21300
ttaggatgca	gtagcagcat	aaacgagata	tggggaagaa	cgagagggtat	ttaaacagtg	21360
atataggccg	ggcgtgggtg	ctcacgcttg	taatcccagc	actttgggag	gctgaggtgg	21420
gcagattacg	aggtcaggag	attgagacca	tcctggctaa	cacggtgaaa	ccccatctct	21480
actaaaaata	caaaaacaaa	attagccggg	cgtggtagtg	ggcgctgtga	atcccagcta	21540
tttgggaggc	tgaggcagga	gagtggcgag	aaccaggag	acagagcttg	cagtgagccg	21600
agatcgcgcc	actgcacccc	agcctggggc	acagagcgag	actctgtctc	aaaaaaaaaa	21660

aaaaaaaaa	attaaaaaa	aaatatttgt	gttaattgtg	atgacaaaa	aaaaaaaaga	21720
gatggaagtc	tctccctaac	ctcactcctc	athtagtgtc	atggcttttt	tctttttttt	21780
tttttttttt	ttgagacaaa	gcctcactct	gtcaccagc	ctggactgca	gtggtgcaat	21840
ctcagctcac	tgcaggtctt	gcctcccaga	ttcaagcaat	tctcctgcct	cagcctcaca	21900
agtagatggg	actacaggca	catggcacca	tgccagccta	attttttgtg	tagtttttag	21960
tagagacagg	gtttcactat	gttggtccagg	ctggtctcga	actcctgagc	tcaagtgatc	22020
ccccctcctt	ggcctcccaa	agtgtctggga	ttacaggcgt	gagccctgct	cccagactcc	22080
tggctttttt	tttttttttt	aatgaaaaat	tcaaatgtct	cttttttttt	tttttttttt	22140
tttgagacag	agtctcgtct	tgctgcccag	gctgaagtgc	agtgggtgtga	tctcggtcca	22200
ctgcaacctc	tgctcccgag	gttgaagtga	ttctcctgcc	tcagcctctc	gaacagctga	22260
gattacaggt	acgcgccacc	acgcctggct	aatttttgta	tttttaatat	agatgggggt	22320
ttgtcatgtt	gtccaggcta	gccttgaact	cctcgctcgc	tgtgatccac	ctgccttgcc	22380
ctcccaaagt	gctaggatta	taggcgtgag	tcgctgcacc	tggccacaga	aattttttga	22440
agaagataaa	taaggtgaca	tttttaaggg	tcaaagaaaa	tgtcaaaaac	tagaatgatg	22500
tctttacata	gggttttaaa	ctttccaaat	taacagggaa	aataattctt	taccttgaaa	22560
ataaatgttt	gctagtgaat	gcaaatacaa	tctttttact	aaatgtttta	ttaaattttt	22620
ttttcttgta	gacacaggg	cctactgtat	taccagggct	ggctctgaat	gcctggcctc	22680
aagcaatcct	cctacctcag	tcttccaaag	tgctgggact	acagacatga	gccatcacac	22740
tgggtctttt	ttaccaaatt	atagtagaaa	gcactttttc	tctaattggtg	aactatgaga	22800
gaattaatca	ggggtctatta	gtaattcatc	cctgaattaa	tcagtgatta	taatgctttg	22860
tgggtcatgt	agtttgctgg	ggattaacac	accatgaaag	tctaccagga	gatttttttt	22920
tttctttgag	aacaggggcc	ataatcagta	gtccttaaat	gaaatggact	attcccattt	22980
cattatatgt	tgcttaggct	ggactcgagc	tccttggtct	aagtgatctt	cccacttcag	23040
actctcaagc	agctgggact	ataggtgtgt	gccaccatac	ctggcttaat	tgagtgtttg	23100
ttttttgttt	ttttttctga	gacagagtct	cgttgtgtcg	cccaggctag	agtgcagtgg	23160
cacaatctcg	gtcactgca	acctctgttt	cccaggttca	aacgattctc	ctgcctcagc	23220
ctcccaagta	gctgggatta	caggcgcttg	ccaccatgcc	tggctaattt	ttgtattttt	23280
agtagagatg	gggtttcacc	atcttggcca	ggctgatctc	gaactcttga	cctcatgatc	23340
cacacacctt	ggcctcccaa	agtgtctggga	ttataggcgt	gagccaccgc	gcccggcttt	23400

aattgagatt	tttagatata	tattactctg	ctaattttgt	cacttgcaag	ttgccatcag	23460
aaaattgtag	gaaaatggat	atatttgttc	cttggaatgg	tttgtgtgag	aataacttaag	23520
gattaaatag	ataagtaaaa	ctggtgggct	ttatataaca	tagatgagca	aatgtcagga	23580
acatacaact	gtgcacacag	ttcaggagaa	ggaggattta	agttaatcaa	caaatttact	23640
aagtataata	aagataactaa	aagtagtggt	tccataccac	tttattactt	aaagtatcat	23700
catataccct	atatttatatg	atatttgcga	caagtcagag	ttaggtaaaa	gaaataacttg	23760
cttttcaggt	aaggagtttg	acgcccagac	agattaactg	acttttccaa	aatcatattg	23820
ctattaaatg	gtggaacaag	gacttaaata	tttgcccttct	aactcacata	cttgcaaaca	23880
catatcctct	cactctaccc	caagctaccc	atgttttgac	ccttcttggtg	gcaatctggg	23940
tctcactaat	atttgaaaga	aaacgtacag	tagataattt	gcaagttaat	ctgttacgca	24000
tatctcttac	ctctatttaa	agatgaatat	cagcatttct	gttgtttcta	cagtaacata	24060
ctaaaaaata	atgcagtgca	ggtgcagtg	ctcacacgtg	taatcctagc	acttttggaa	24120
gttgaggcag	gaggatcact	tgaagccagg	agttcgagac	tagcctgggt	atgcaagacc	24180
ccatttctgt	tttttttttt	tgtttgtttt	gttttgtttt	gttttgtttt	tttgagacgg	24240
agtctcgctc	tgtcgccag	gctggagtgc	agtggcgga	tctcggtca	ctgcaagctc	24300
cgctcccgg	gttcacgcca	ttctcctgcc	tcagcctccc	aagtagctgg	gactacaggc	24360
gcccgcact	acgcccggct	aattttttgt	attttttagta	gagacgggt	ttcacctgtt	24420
tagccgggat	ggtctcgatc	tcctgacctc	gtgatcccca	tttctttttc	actgcaacct	24480
ttgcctccca	ggttcaagt	attgtcgtgc	ctcagcctcc	caagtagctt	gggattacag	24540
gcatgtgcca	ccatgcctgg	ctaatttttg	tattttttgg	tagagatgga	gtttctggcc	24600
aagctggcca	ggctggtctc	gaacgcctag	cctcaagtga	tctacctcc	ttagcctccc	24660
aaagtgctgg	gattacaggt	gtgagccact	gtgcccggcc	ccatttctac	aacaattaaa	24720
aaatattagc	ccagtgtagt	ggtgcatgtc	tctagtccca	gctactcaga	aggctgaagt	24780
gaaaggattg	cttgagccca	gaatttcaag	gctacagtga	gctatgataa	tggcattgca	24840
ctccagcttg	ggtgacagag	tgagaccctg	tctctaaaaa	atgaagtaaa	atagtgcaca	24900
agtatagaac	ttgaaaatct	tccttaacct	taccataagg	gaaatgatta	ctaataagtt	24960
tcttaacttt	ttgtacttac	ataaacataa	atattcatca	gagaaaaaaa	tatgcaaaac	25020
aatttgcaat	ctttttcact	taccatattt	tggaaattttt	ttcatttcaa	tatatttgat	25080

cttccttggtg	ttttttcagt	ttgtttgttt	ttgtcaccga	ggttgagtg	aagtggtaag	25140
aacatggctt	attgcagcct	caatctcctg	ctcccattca	gcccctcaag	tagctgggac	25200
tacaggtaca	tgtcaccacg	cccggctaata	ttttattttt	attttggtag	agatgggggtt	25260
tcaccatgtt	gcccaggctg	atcttgaatt	cctgggctca	agtgaccgc	ccacctcagc	25320
ctcccgaagt	gttgagagaa	caggtgtgaa	ccaccatgct	ccacctctta	gtctttacaa	25380
tctgcaaaac	ctcataagtg	gctaatagag	gaatatagta	aagcaaagg	ggatatcact	25440
gattagaact	gtgttttttag	gctgggtgcg	gtggctcag	cctgtaatcg	caacattttg	25500
ggaggctaag	tgggagtatc	acttgagccc	gggagttcaa	gaccagcctg	ggcaatatag	25560
tgagagaccc	tgtctttaga	aaaaaattaa	ccaggtgtgg	tgggtgcacac	ctgtgggtccc	25620
agctattcaa	gaggetgagg	tgggagaatc	gcttaagcct	aggaggcgga	ggttgcagtg	25680
agatcatacc	actgctctct	agcatgggtg	acagagcgag	accagtccta	aaaaaaaaag	25740
tattttttcg	tttttttcca	actcatgtac	accgccacc	ccaccctgc	tttttttttt	25800
tctgacattg	ggtcttgctc	tgtcaccag	gctagagtgc	agtagcaca	tcaactcact	25860
gcagcctcct	cttcctggac	tcaagcagtc	ctaccacctc	agcctcccaa	gtagctggga	25920
ccacaggtgt	gcaccatcat	gcctggctaa	tttttgtact	ttttgtagag	atagggtttc	25980
accatgtggt	tcaccatctc	taactcctgg	gctcagtcag	tccacttttg	cctcggcagtg	26040
agccactgtg	tgcagccac	gttttttatt	aatggatatt	tggattgttt	ccatctattg	26100
tgaataatgt	ggctatgaac	attgggtctaa	atatctgttt	aagtcccggc	tttcaatact	26160
tttgatata	tacctaggag	tagaattact	gaattatatg	gtaactttct	gtttaaattt	26220
ttgaactgcc	aacctgtttt	ccataggggc	tgcaccattt	tgcattccca	ccagcagtg	26280
acaagggttc	cagtttctcc	acatttggtta	tttttcattt	tttaaaataa	tagtcatcct	26340
aaagggtatg	aagtggatc	tcgttgtgat	tttgatttgc	atgtattttt	ctaatactt	26400
atgatgctca	gcattttgtc	atgtacttat	gtaccatttg	tgtatcttct	ttggaaaaat	26460
gtctattaat	gttcttttcc	cattttttta	ttgggtgtgt	tttatgttta	tcaattttgt	26520
aaacatttta	agctctgcag	cataactact	caaccctgtc	acatggtaag	attgaccag	26580
taaaacttta	tgtacaaaaa	taggcagctt	actagattta	atcttagtcc	atagtttgct	26640
aaggcatgca	ttagataatg	tagttacact	attggctaata	aatttaaaact	acaagtgggtt	26700
gtaagtttct	gccacccaaa	ttctttcttg	atgtgatgta	gtctgggttg	ttgaatttga	26760
gtgttatatg	gggtcataga	gttaaaagag	aaatgtctat	gagaaactag	ggactgttgg	26820

gagctaattgt	taaaggatttt	tggaggcctt	ttgtgcactg	gagaccattg	gaagattgga	26880
ttttctgcta	taaatgtatc	taaaggataa	tcagtgtaa	ttatgggctg	tagtttgcca	26940
acccctacat	taagggattg	aaataatttg	aatatgggtt	tcagttcttg	tatggctctg	27000
ctcagttctt	tttttttttt	tttttttttt	tttttttttg	agatggagtc	tcactctgtt	27060
gcccaagctg	cagtgccatg	gcacgatctc	agctcgctgc	aacctccgcc	tcctggcttc	27120
aagtgattct	cctgcctcag	cctccagagt	agctaggact	acaggcacgt	gccaccacat	27180
ctggctaatt	tttttttttt	tcaagacaga	gtttcgctct	tgtcaccocag	cctggagtg	27240
aatggcacga	tcttggtctc	ccgcagcctc	cgctcctgg	gttcaagtga	ttctcctgcc	27300
tcagcctccc	aaatagctga	ggcaacaggc	gtgcgccacc	acgcctagct	aatttttctg	27360
tttttagtag	agatgggggt	tcaccatctt	ggccaggctg	gtctcgaact	cctgatctca	27420
tgatccaccc	gcctcagcct	cccaaagtgc	tgggattaca	ggcgtgagcc	acctcacctg	27480
gcgttttttt	gtgtgtgttt	tcagtagaga	tggggcttca	ctgtattagc	cagaatggtc	27540
ttgatctcct	gacctcgtga	ttcacctgcc	tcagcctccc	aaagtgcctg	aattataggc	27600
atgagccacc	acgcctggct	gtggctcagt	tcttaactgt	tcattcattc	agggtcctag	27660
ccaaaatctt	ttgccacagc	ttctcctggg	catgctctga	aattcatttt	ttgtctgttt	27720
agttgcatga	gtccactgaa	agctttgctt	attttctcca	ccactttcag	agttcacagt	27780
cacttcaagt	agacgaggtc	aggagatcga	gaccatcctg	gccaacatgg	tgaaccccg	27840
tctctactaa	aaatacaaaa	cttagctggg	cgtggtagcg	tgtgcctgta	gtcccagcta	27900
ctcaggaggc	tgaggcagga	aaattgcttg	aaccgcagag	gcagaggttg	cagtaagccg	27960
agatcatacc	atttcactcc	agcctgggcg	acagagtgag	acactgtcaa	aaaaaaaaaa	28020
aaaaaaaaaaa	aggaaagaaa	agtgtcccaa	atgctgggct	catctctgct	ggcttctctc	28080
ctttttatat	ttcgataatt	tttaaaataa	taaccatcct	aatgggcttg	atatgggtccc	28140
tcctttctaa	attttggtaa	ttttcattgt	tttggaact	cactaatccc	ttcaaacaga	28200
tgtttttaaa	actttttttc	tagttctctc	tctttttttt	tttttttttt	tttttgagac	28260
agggctctct	tcagtgggtg	gacctccctt	actgcaacct	ctccctcctg	ggttcaagca	28320
gcgattctcc	tgcctcagcc	tcccaaatag	ctgggagtac	aggcaccac	ccccacgccc	28380
agctaattgt	tgtattttta	gtagagattg	ggtttcatca	tgttgccag	gctggctctg	28440
aactcctgac	cccaggtgat	ccgcctgcct	cagccttcca	aagtgcctgg	attacagcca	28500

tgagccaccg	cacctggctc	tagttctctt	ctaattcttac	tatagtaatc	aaaatttttag	28560
tgctgatgtt	ataatccaag	cccaaattgga	tattctttata	cattaaatgt	tggaatatca	28620
tgtttggttaa	aaatcagatc	tgctgggcac	ggtgggtcaa	tgctgttat	cccagcactt	28680
tgggaggccg	aggcaggtgg	atccccctgag	atcaggaggt	cgagaccagc	ctgaccaaca	28740
tggagaaact	cogtctctac	taaaaataca	aaattagccg	ggcgtggtgg	cacgtgcctt	28800
tcatcccagc	tacttgggag	gctgaggcag	aagaattgct	tgaacctgtg	aggcggagggt	28860
tccagggagc	tgagatcgtg	ccattgtact	ccagcctggg	caacaagagc	gaaattccat	28920
ctcaaaaaaa	aaaaatcaga	tctgtcccta	tggtttttatc	ttcactgcat	gtcatatata	28980
aatggaagca	tgtagtatgt	agccttttgt	gtctcgcttc	tctccctttg	cataatattt	29040
ttgagattta	tcccttttcc	tacatatgtc	agtagtttgt	ttcttttttaa	tgctgaatag	29100
tattccacgt	gtggagatac	cacaatttgt	tactccattc	actaaatgat	ttgggttgtt	29160
tcctgttttt	gttattgttg	ttgttgttgt	ttaattaatt	aattaatttt	ttttttgaga	29220
tggagtcttg	ctctgttgcc	caggctagag	tgcagtggcg	caatcacagc	tcactgcaac	29280
ctctgcctcc	coggttcaag	tgattctcct	gcctcagcct	cccaagtagc	tgtgattaca	29340
ggtgcccgcc	accacaacca	gctaattttt	gtattttttag	tagagatggg	atttcaccat	29400
gttggccagg	ctgggtcttga	actcttccct	ccttggtctc	ccaaaatgct	gggattacag	29460
gtgtgagcca	ccatgcccac	cctcctgttt	ttaaattttta	tgaataaagg	tactgtaaat	29520
tcatgtacag	gtcttagaag	tttgcatttt	tcttgggtaa	atatgtagaa	gaagagattg	29580
ttgactcatg	tggtaaatat	atgtttaatg	tcataagaaa	ctaccaaact	gtttttccaa	29640
ctgggtgcca	ttttgtttcc	taccagcaaa	cataatgaaa	ggtccattac	cctttgtctc	29700
gtaatatattt	ggatattttta	tgttcttttg	tttttttgcc	attcaaaatt	gggtgcattg	29760
ttgtatcttt	ggattttaaat	ttgtacttcc	ctaacaacta	atgatgtgga	gcatcttttt	29820
atagccttac	ttgccaatta	tgtcttaatt	ctttggcaat	acttatattt	gactctgcct	29880
actatatata	tatatatata	tatttttatg	ctatatatag	agtatatata	tttttatact	29940
atatatatatt	atatatacta	tatatattta	tatatactat	atatatatat	atatatatat	30000
atattttttt	ttttttttga	gatgaagtca	tgctctgttg	accaggctgg	atgcagtggc	30060
atgatcttgg	cttaccgcaa	cctcaacctc	tgctcccag	gttcaagtga	ttctcctgcc	30120
tcagcttcct	gagtagctgg	gactacagac	acatgccacc	atgcccggt	aatttttgta	30180
tttttagtag	aggtggagtt	tcgccatgtt	gtccaggctg	gtcttcaact	cctgacctca	30240

ggtgatccac	ccgcttcagc	ctcccaaaag	tgctgggatt	acaggcatga	cccaccgtgc	30300
ccagctggaa	taagcaatct	taaaaagcag	tttgttgatt	tctggtgaat	agagaaaatg	30360
tgacaaacta	tgactcaata	gtttcatttc	ttttttgttt	gtttgtttgt	ttttttgaga	30420
cagagtcttg	ctctgtcacc	caggctggcg	tgcagtggct	cactgcaagc	tctgcttcct	30480
gggttcacgc	cattctcctg	cctcagcctc	cggagtagct	gggactacag	gcgcccgcga	30540
ccatgcctgg	ctaatttggt	ttttgtattt	ttagtagaga	cagggtttca	ctgtgttagc	30600
caggatggtc	tcgatctcct	gacctcgtga	tccacccgtc	tcggcctccc	aaagtgctgg	30660
gattacaggc	atgagccacc	gcccctggcc	tacttttatt	tcttaagcat	atatcttaaa	30720
gaaactgcag	gtatgcatta	ggggacaggt	agaaaaatgt	tacagcattg	tttgaaatgt	30780
tgaaaaactg	gaaacaacct	caatatctta	ataatagaaa	aaatggttga	ggctggctgt	30840
ggtggttcat	gcctgtaatc	ctagcacttt	tggaggtga	ggcgggcgga	tcacctgagg	30900
tcgggagttt	gagtccagcc	tggccaacat	ggtgaaaccc	cgtctctact	aaaaatacaa	30960
aaaattaggc	aggtgtggtg	gtgcacacct	gtaatcccag	ctacttggga	ggctgaggca	31020
caagaatcac	ttgagcctgg	gaaatggagg	ttgtagtgag	ccgagatcat	gcctactgca	31080
ctccagcctg	ggcaatagag	tgagactccg	tctcaaaaaa	aaaaaaaaaa	aaaaaaaggg	31140
aaggagcatg	ggaatgggaa	gaggcactaa	aaaaggaact	tcttttgtat	ctgttacatt	31200
gtactcattt	tgtttttaaa	attttaagcc	aaatatgaca	ttgttatcac	ttttgttgcc	31260
actcagaata	caagtgctat	tacactattt	tctataattt	tccctttttc	aaactaaaaa	31320
acaaaaatac	atgtaaacag	gtgtttttctt	tatggaaaac	aggaaggtga	gcatccaaat	31380
aatttttagtt	aattttataac	aaactcttaa	taattttcttt	ttcttcttct	totttttttt	31440
tttttttttt	tgagtcaggg	tcttgctctt	ttgctcaggt	tagagtgcaa	tgggtacaatc	31500
tgggtacaatc	acatcttagt	gcaacttcta	cctcttcagc	tcaagtgacc	tctctcctca	31560
gtcccccaag	tagatgggac	cacagatgtg	tatccccaat	catggttaat	ttttaaattt	31620
tttggtagag	ataaaaatctc	accatgttgc	ccaagctggg	cttgagctcc	tgggctcaag	31680
cagtctcctc	gccttggcct	cccaaagtgc	tgggattata	ggcgtgagca	aactgtgccc	31740
agtcttcata	aacttttctt	tacatgtcct	tatcaagtac	tttttgagca	cctactgtca	31800
accagccgta	gtatataata	ctgataattc	tataacataa	gaaattgacc	tgtttaaggg	31860
gatgaatatg	gttagttatt	ttcagggtga	agaaacagaa	gaatcgggga	ggtagtagat	31920

agtcataagg	agacggcatt	tcttagtcac	tttgtgtggt	gtttataata	agacttcattg	31980
tggccagggtg	cgggtggctcg	tgagcgccctg	taatcccagc	actttgggag	gccgagggtgg	32040
gaggatcaca	acaaggtcag	gagatcgaga	ctatcctgaa	caacatggtg	aaaccccgtc	32100
tctactaaaa	atacaaaaat	tagctgggtg	tggtggcacg	tgccctgtaac	cccagctact	32160
cgggagggtg	aggcaggaga	atcgcttgaa	ccagggagtc	ggaggttgca	gtgaaccaag	32220
atcgtgccac	tgactgacg	cctgggctac	agagagagac	tccatctaaa	taaataaata	32280
aataaataaa	taaatagact	tcgtctgttt	catcaagagt	cattgtatta	tattgatttt	32340
gaaatttcaa	ataataataa	ggaatagtat	ggaaattttt	gtttgattaa	atggggacat	32400
gaggccagggt	gtggtcactc	aggcttgtaa	tcccagcact	tcgggagacc	agggcaggag	32460
gattacttga	gcctaggagt	tcaagaccag	cctgggcaac	atggcaagac	cctatctcta	32520
caaaagaaat	acacaaaaat	tagctaagtg	tggtggcaca	tgccctgtagt	cttagctact	32580
cgggaggcca	agggaagagg	atggatcact	tgagcccagg	agttggaggg	tgcaagtgcg	32640
taggattgtg	cccctgcact	ctagtgtggg	tgacagcaag	actctgttta	aaaaaaaaaa	32700
aaaatgaggc	cgggcgcggt	ggctcacctc	tgtaatctga	gcactttgag	aggccaaggc	32760
gggtggatca	cctgagttca	agatcagcct	ggccaacatg	gtgaaaaccc	cgtctttata	32820
cttaaattac	aaagattagc	tgggtgtgat	ggtgcacacc	tgtaacccca	gctactccag	32880
aggctgagac	agcagaatcg	cttgaacccg	ggagggggag	gttgcaatga	gccgagattg	32940
caccactgca	ctccagccta	ggtgacagag	caagactctg	tctcaaaaaa	aaaagaaaaa	33000
aagtctaaga	agttaatttt	cattcagaca	aatgttcaaa	aataatagaa	ataaaaacaa	33060
aaattaattt	tcaaatatgt	gatttgagta	taacgatcac	tttacagaag	ttattctaca	33120
tactttttga	ttatgctgag	tttttaggtc	cttttaaagc	ttaacttgta	tcaggaatta	33180
atgtttttaa	ctagttgttg	ggttcaaaga	aaatagaaat	gtgttattaa	atgccacact	33240
tgtaatctca	tacagcagtt	ctcatgatgc	aatcaattgt	taaataaaaa	cttcctctag	33300
agaggaggta	cccagcaatg	aacacattat	atcagatttg	aatatgagat	taaacaatgc	33360
tttttgtttc	attgttattg	aaatcatatt	tgtatttttc	aaaagtatat	atacttaaaa	33420
attcctaaat	aaaatgtctt	tctgtaatac	agcatttttag	gttataagga	tcaataccat	33480
gggctgataa	gcaaaatagg	actcacggcc	ccaaaatggt	gataaacatc	atgaccatat	33540
tccccataat	ggagttacat	atacacatta	caatgataaa	aataactaaga	agttatatag	33600
taatgtttct	cagacttggt	cacaaccctg	ttttttatgt	attacctctc	aaaagcatga	33660

tttgcaatat	ttagtgtact	cttcatacga	tctatacaaa	taatttttta	ttttaaagca	33720
ggacagaagg	agcaagctgt	ggaatggtat	aagaaaggta	ttgaagaact	ggaaaaagga	33780
atagctgtta	tagttacagg	acaaggtaag	attgtatttg	tttatagcca	tcccaaatta	33840
tgatatattc	acatgattgt	ccagatttca	gatctattta	tttatttatt	tatttttctt	33900
tctttctttt	cttttctttt	cttttttttt	tttttttttg	agacagggtc	tttcctctgt	33960
taccagggt	ggagtgcagt	gatgtgatca	tagctcactg	caacctcaaa	cttctgggct	34020
caagtgatcc	tctgcctct	gcttcccatg	tagctgggac	cacaggcgca	cactaccata	34080
cttagcta	tattattatt	attattatta	ttattattac	tagtttttga	gacagagtct	34140
cgctcttttg	cccaggctgt	agtgcaatgg	tgcaatctcg	gctcactgca	atctccgct	34200
cccgattca	ggctatttat	tctcctgcct	gagactcctg	agtagctggg	attacagggtg	34260
catgccaccg	tgcccagcta	atttttgtat	tttttagtgaa	gatgggggtt	tgccacgttg	34320
gccaggctgg	tttcgaactc	ctgatctcaa	gtgatctgcc	cacctcagcc	tcccaaagtg	34380
gtgggagtac	aggcatgagc	cactgtgggc	cacacttagc	taattaaaaa	aaattttttt	34440
gtagagacag	ggtctcactg	tgtttctgtg	actgatctta	aacttctggc	ctcaagtgat	34500
cctcctgcct	cagctttcca	aagttgtggg	attacaggca	tgagccactg	taccagcca	34560
gattgtattt	atcttaataa	atgtatttgg	gccaggcaca	gtgggtcatg	cctgtaatcc	34620
cagcactttg	ggaggccgag	gtgggcgaat	cacctgggat	caggagtctg	agaccagcct	34680
gatcaacatg	gagaaaccct	gtctctacta	aaaattcaaa	attagccagg	cgtggtgtcg	34740
cgcacctgta	attgcagcta	cttgggaggc	tgaggcagga	gaattgcttg	gaccaggag	34800
gcggagggtg	tggttaagcca	agatcgcgcc	attgcactcc	agcctgggca	acaagagtga	34860
aactctgtct	cagaaagaaa	gaaaagcaaa	acagtatttc	atttaattaa	aatgccttg	34920
tattatacca	gtaagagtaa	aaaacaattt	aattaaagta	tgatatatgc	tcttttcac	34980
acttagaatt	tgaattttta	tcttattgaa	gaggctgttt	tagttttatt	tagatatggg	35040
tataacaaca	atattaatat	ctgtgaccaa	attggtgcat	gcatgggcaa	tgaaaagcta	35100
tgttgtgact	gcagcatggg	ggatagcatt	agtaagttag	cattgggtgt	aaaatgaatc	35160
ccaatttcaa	agattttgaa	acgtggggaa	aaaatatttt	agaccccatg	aaataagacc	35220
tgaaatctat	gaaataccac	agtaggttgg	aatcatcat	gagaaactgt	aactattttt	35280
ttttataggt	attatatgta	ttctttgaga	cagggtctca	ccctgttgcc	caggctggaa	35340

tgctgtgatg	tggtcacctc	tactgcaac	atccacttcc	tggggctgaa	gtgatccttg	35400
caactcaccc	tcccatgctc	ggctaatttt	tttttttttt	tttttaagta	gagaccagat	35460
ctctttatgt	tgcccaggct	ggtcttgaac	tcctgggctc	aagtgattct	gctttgtctg	35520
cccagagtgc	tgggattaca	gaagtgagcc	acaacacctg	gcctaaagat	agtattttta	35580
aaatttcttc	tgttttcttt	cttttgggta	tacattttct	tctttttttt	aaaaaaaaat	35640
ttctgtataa	agactgtgac	tcccatgaa	agtagtttgg	gtgataattt	atcgtgaaac	35700
aattattagt	gggaaatgta	gatattttta	ttaatttttt	tctttcagg	gaacagtgtg	35760
aaagagctag	acgccttcaa	gctaaaatga	tgactaattt	ggttatggcc	aaggaccgct	35820
tacaacttct	aggtatcaat	taatgtataa	tttgatgtgg	gatgtattgg	aaatgtgtgt	35880
tcaatgaaac	tttaatttgt	agaaagaaat	agatcagtga	ttgaaaatgt	gggccaggct	35940
ttttaacata	aagaaaacgt	ataacatata	caaaaacaga	agagcataat	ggacttcttt	36000
ttttttttga	gatggacgct	tgtctgtca	cccaggctgg	agtgcagtgg	cgtgatcttg	36060
gtcactgca	agctctacct	cccaggttca	cgccattctc	ctgcctcagc	ctcctgagta	36120
gctgggacta	caggcacccg	ccaccacgcc	ccgctaattt	ttttgtagtt	ttagtagaga	36180
cagggtttca	ccatgttagc	cagaatggtg	tccatctcct	gacctcgtga	tccacccgcc	36240
tgggcctccc	gaagtgtctg	agttacaggc	gtgagccaac	gcgccagcc	gagcataatg	36300
aaattcta	tattcttctc	tcagctttga	taattatcac	ctcactatca	gtcttgattt	36360
atttatacca	agggttagca	aactgcagct	gatgggcca	cccagctcac	tgcctattac	36420
tataaataag	gttttattgg	cacacagcca	ttcctttcct	ttaggtattg	tctatggctt	36480
tcattgtacg	gtgacagaat	tgaatagttg	atacagagac	aacctgccc	tcaaaacct	36540
acataatttac	tgtctggccc	tgagaaagcc	tgccatcttc	acttgacatc	ccacttact	36600
gtcccccaac	ttcttattat	tttgaaacaa	attattgact	tattttccat	ttataaatat	36660
ttcatattat	acctctaaaa	gataaggatg	tcaaaataaa	atcacagaaa	taccattatc	36720
acacctaaaa	aataactcca	aatttctttt	aacattatcc	aaaaaaactg	accccccaa	36780
aatgcaattt	ccaaatttct	aattttctca	aaatttaata	atttttacta	ttttttaata	36840
atttgtttgt	atcaggatct	taaaaagatc	cgtatgtctc	ttttaatgta	caggttcatc	36900
tgtttttcat	tttcttccaa	tttatttggt	gatgaaacca	ggtctttgtc	atgtaatatt	36960
tcctacagtc	tgacttttgc	tgcattgcac	ttttgtggta	tgttttaaca	tgttcttctt	37020
tcctccacg	ttaaatgata	gttgatata	gagccttgat	cacattgaaa	gttgattttt	37080

atTTTTtaat	gattagacta	cttCctaggt	ggtTgtgttt	gtgtgttcat	cattagttgt	37140
tccatcttgt	gaggttagca	gcagttaact	atgactgatg	tccagtgata	ctttaattct	37200
gtcattcttt	cagcattcat	caactgtaat	agaaactttt	acttgtcttc	tgtttgatgg	37260
cgtaggtttt	aaatatatat	taataagtta	aatctaccac	tcaaaaaaag	tagaacaaaa	37320
cctactacat	atgagtggct	ctaccatacg	ttggtggatg	gaaatagtag	atatttggtta	37380
agcaggtaaa	ccttattttat	ttgtagccct	gtgacttggg	gtaagttatt	ttgcatgtct	37440
gggtcttttt	tatctaattt	gtaatatgaa	gctattatta	ttattattat	tattattatt	37500
attattatta	ttattattat	tattcgagat	ggagtcttgc	tttgttgccc	aggctggagt	37560
gcagtgggtc	aatctcggtc	caatgcaccc	tccacctcct	gggttcaaac	tattctcgtg	37620
attcagtcct	cagagtagct	aggattacag	gggcctacca	ccacacctaa	tttttgtatt	37680
tttagtaggg	acgggggttc	accatgttgg	ccaggctggg	ctcaaactcc	tgacatcaag	37740
tgatccatcc	acctcagcct	cccaaagtgc	tgggataggc	cgggcacggg	ggctaacgcc	37800
tgtaatccca	gcactttggg	aggccgaggc	gggtggatca	cgaggtcagg	agattgagac	37860
catcctggct	aacacggtga	aaccccgctc	ctagtaaaaa	aatacaaaaa	actagccagg	37920
tgtggtggcg	ggcacctgta	gtcccagcta	cttggggaggc	tgaggcagga	gaatggcatg	37980
aaccacaggag	gcggagcttg	cagtgagccg	agatcatgcc	actgcactcc	aggctgggca	38040
aaagagcgag	actctgtctc	aaaaaaataa	aataaaataa	atagataaat	aataataata	38100
aagtgtctggg	attacaggca	tgagccacca	cgcccagctg	aagctaatat	tattagctaa	38160
taataatatc	tcatattttt	gaccattgaa	ttattaaagg	aacctaaacc	atagtagtaa	38220
gtgctttaaa	aatgttggtt	tttattttatc	tgttttacctg	actgtcatca	cctcctttgc	38280
ttaggtagaa	agatatattc	ataaacacat	acatatacat	tttatttcatt	ctttaattca	38340
tttactccta	tatgggtcatt	gtgttttgtgg	gattttttgcc	cgcacatagt	taaaaatgca	38400
gccgttactg	ctttcagatg	agtaacaagg	tagtgttccc	tggcttctgt	ggctgacagg	38460
at ttgccttc	ttcctttcta	aatggagggtt	attactgtgt	cagatataat	taaatagtgt	38520
attaaagttt	gttgcaataa	aataattgat	ggttctaatt	ggtactttct	acgtgtttta	38580
tctttaaagc	ttttcagtgt	atatataaag	tatatatcat	acagaataaa	tttgattgtg	38640
gagcattttg	taacatcttt	caaaattaat	aagagttggg	ttttatgttt	tgtttgatgt	38700
tatgatatag	gacattagaa	gtattaatat	caattagaga	ctcatctttg	aatgtgactt	38760

tgtactttct	tatttgtgtt	agtagaggag	agaacaaaaa	gaagatatgt	aatgtaatat	38820
gaaactaggc	atttaaagat	ttaacgtttt	ggatatttta	aagttggtgt	ctgttttcac	38880
cctcaaaaat	gctattacct	atttatgaaa	tatcttttaa	agtgtggagt	ggtagatggg	38940
aaaggtgaca	tcggtctggc	gcggtggctc	acacctataa	tcccagcact	ttgggaggcc	39000
aaggtgggcg	gatcacctga	ggtcaggagt	tcacgaccag	tctggtaaca	tggtgaaacc	39060
taatctctac	taaaaataca	aaaaattagc	caggctggtg	gtacgcacct	gtaatccag	39120
ctactcagga	ggctgagtca	ggaggatggc	ttgaaccacg	aaaatggagg	ttgcattgag	39180
ccaagatggc	accactgcac	tccagactgg	gtgacagaga	gagactccgt	ctcaaaaaaa	39240
aaaaaacgcc	gggcactgtg	gctcacacct	gtaatcctag	cactttggga	gactgaggcg	39300
agcagatcac	ttgaggtcag	gagttcgaga	ccagctgggt	gcctgggcaa	catagtgaaa	39360
ccccgtctct	actaaaaata	caaaaaaatt	agccgggcgt	agtggcgggc	gcctgtagtc	39420
ccagctactt	gggaggctga	ggaaggagaa	tggcgtgaac	ccgggaggcg	gagcttgacg	39480
tgagccgaga	tcccgccact	gcactccagc	ctgggtgaca	gagcgagact	ccgtctcaaa	39540
aaaaaaaaaa	aaaaaaaaaa	tagccgggca	tggtggcggg	cacctctaata	ctcacctact	39600
tgggagggtg	aggcatgaga	attgctcgag	cctgggaggc	agaggttgca	gtaagcagag	39660
accacaccac	tgcactctag	cctggatgac	agagtggagc	tttgtctcaa	aacaaaaaca	39720
aacaaaccaa	accacctctt	gtccactcct	aatctccctt	actgtgtttc	acttattttt	39780
ttcccatata	tttgtcacct	tctaataaac	tgtgtaactt	attttttatg	tttatttgtgt	39840
actatctgtc	ttctccctgt	tggaatataa	actctatgga	ggagggatcc	ttgtctatat	39900
cccaagtgcc	tacaacagtg	tctagcacat	gctagttacc	gaataaatat	ttgccaggtt	39960
aatgaagttg	tttataaccc	ttggcactat	acctgatata	gtgtagggac	ttaataaatg	40020
aatgttaaata	gaagctatct	taaaaatata	tatatatata	taacagccag	atacgccaat	40080
ttgccgacat	caaaattaat	cttttcatct	attgcagggt	ttacagttaa	gctttgctct	40140
aaggttcact	atcactaaac	caaaaataaa	taatattgat	ggtcttttcc	catttttaac	40200
agagaggcta	catcctctca	aaccagtcag	tattaggaat	tcccaggatc	attctctcca	40260
ttttccatcc	taattagata	cctatctggc	accacatctc	tgaaaagctc	caatttttaa	40320
ccagcacata	ataaacaat	atgttactat	taatacacgt	ttgtacacaa	gtcattttta	40380
ttgtcttagg	tatatccact	cccagacctt	ctgagttatt	gaaagacagg	aatttgtgat	40440
gtcacctctt	ttttcttttt	tttgagacgg	agtcttgcc	tatcaccag	gccagagtgc	40500

aatggcgtga	tctcggctca	ctgcaacctc	tgcctcccgg	gttcaagcga	ttatcctgcc	40560
tcatcctcct	gagtagctgg	gattacaggt	ggccctagtt	tgggtttttg	gttttaaacaa	40620
ggatgggtctc	ctctttgaga	agttgccatg	agttagccat	gatgatacct	gggacaataa	40680
tgggtccacag	agaagtaaga	gccaatacaa	cgccttaaga	gaaagtacct	ggtatgtttg	40740
tgaagcagca	aagaggccaa	tgtgttccag	caaggagatt	atagtagaga	tgaagtcaga	40800
gaactcatgc	ggagagggat	tgtcacaaca	cgcagggctg	taaaggccac	catgaggact	40860
tgagcttttta	ccgcctgtga	aatggggagc	agttacaaag	atttttaaga	aagcagccca	40920
gtgcagtggc	tcatgcctat	aatcccagct	ctttgggagg	ctgaggtggg	tggatcacct	40980
gaggtcagga	gttcgagacc	agcctggcca	acatggcgaa	accttgtctt	tactaaaaat	41040
agaaaaatta	gctgcgtgtg	gtggccatgc	ttgtaatccc	aaatactcgg	gaggctgagg	41100
cagaagaatc	acttgaaccc	gggaggcgga	ggttgcagtg	agccgaggtt	gcgccactgc	41160
cctccagcct	gggcaacaaa	gcgagactcc	gtctcaaagc	cagaaaaaaaa	gatttttagg	41220
aaaggaatga	aataactctga	cttaaatttt	agaaggatca	ctctggttgc	tatgttgcac	41280
tatagactct	aagggtagaa	ccagggagag	caattaggac	actgttacag	taatctagga	41340
gaaaagtgat	agaggcttgc	accagggtag	tatcaggaga	aatggtaaca	agtggttgga	41400
ttatgattat	agttgaaggt	cgaacaacta	gatttgctga	tagattggat	aaaatggtcc	41460
tctctttatg	gatctcataa	tacagtagga	gaaacagaca	cataaatagt	catcactgca	41520
ggatgggtat	tagaaatatt	ttgcaagggt	gtaaaagggtg	ctaagaaagc	tttcttgtgt	41580
gacaaagctg	gttaaagaaa	gaaaagaaaag	cttcctagaa	gtgaaaccta	aaatatacac	41640
ttagataaat	gaataaaaagt	tatgaaagaa	tgtgagggct	gggcacagtg	gctcatacct	41700
gtaatcctag	cactttggga	ggttgagtcg	ggggatttgc	ttgagtccag	gagtttgaaa	41760
ccagcctgga	cagcgtactg	agaccccatg	tctaaattta	aaaaaaaaaaa	aaatagccag	41820
gagtgggtgat	gtgcacctgt	agtcctagct	acttacttgg	gaggctgagg	caggaggatt	41880
gcttgagctc	aggaggtcga	gactgcagtg	agccatgagt	acaccactgc	actctagtct	41940
gggtgacaga	gtgagaccct	gtctcaaaga	aaatgaaaag	aaaaagaatt	tgagaaaaga	42000
tatgtcagga	aaatgtgatg	acatgaacaa	atactcaaag	gcaagaaaag	catggtgagt	42060
aggggaggggt	acaggttgag	acatgaagtt	gggattgccg	aggcgtaaga	ggtataggca	42120
gatcatggat	ggcctggttt	gtcataggaa	ggagcttggg	ctcttatctg	tgggcaatgg	42180

gaagccacta	aagggtttta	agtagaagag	tgttatatgg	taagggttttc	ctctccagta	42240
gatcactcag	atgactgtgg	gggttggatg	tgaggaggtta	aagcagcaaa	gttactgctt	42300
taattttaaat	cattgattaa	tctgtacaat	cctagtcatt	ccaaaaagaa	acatttagtcc	42360
tttaacagaa	gtgaattggg	tctgggcgcg	gtggctcacg	cctgtaatcc	cagtactttg	42420
ggaggccgag	gcgggcagat	cacctgaggt	caggagtttg	agacaagtct	ggccaacatg	42480
gcaaaacctt	gtatctacta	aaattacaaa	aattagctgg	gcatgggggc	acccacctgt	42540
aatcccagct	attcgggagg	ctgaggcagg	agactctctt	gaacccataa	ggcagagggt	42600
gcaatgagct	gaggtcgcac	cactgcgctc	cagcctgggc	gacagagtga	gactccgtct	42660
ccaaaaaaaa	gaaaaaatat	tggagcagtt	tcacagatgc	tgtttactgt	tatactgtat	42720
gtgtctatga	ctcctcctcc	aagaaaaaaaa	aaatgaattg	gagcaggttc	acagaagcaa	42780
tgtactgtta	cactgtatgt	gtctgtttct	acacatataa	atctgaattc	tgtgtacacg	42840
aaaagaaatc	ctatagcttt	ttattcctag	ctataaaaac	taagaatata	atttctttct	42900
ttttattaca	taatgaatgg	ttctgttaac	tttttgttaa	aatttcattg	aggggagtat	42960
taattcacat	ttatgtacgg	atgtctacat	ttacaaatca	gtgtattttt	tgattttatgc	43020
ttttactgag	acaaaggggt	tctgtctcag	catggtcatt	taaagagttt	atcattgaga	43080
aaaatcagat	gaccaacctg	ttagctcaaa	aaaaaaaaaaa	cctccaaggt	atattgtatc	43140
agccagttct	aggatacaaa	agccatgcag	tacttttgtgt	tttgtgccaa	aaagggtagc	43200
tgctatttga	cctgtcccaa	aggcatgtgt	ggttgtagcg	taaaccaagc	atggtacctg	43260
tttgtcaaac	tttagaaatg	aaagtttaag	agagttaata	tataggtgct	gcatttttta	43320
tgtattcatt	gacttgctgg	tacagaagaa	aagaatcaat	tatgattcag	cacaatactc	43380
cacttgggga	agagagtgca	gcagtagttt	agagtgtcag	ggatcaaact	gctaccttct	43440
tgggcttcag	ttgctagact	taagagaccc	agatcttggg	aggggttttt	gttgttattt	43500
gatgtggggg	taaggttgaa	atactccctt	ttaattgatt	aatataaaaa	ataggtaatg	43560
tagacgtatg	tcacaaaatt	ttaacaagtt	tgagtgtgaa	aagcatccat	cccattctat	43620
gcoctggcta	cccatttccc	ctccctagag	goggetgtta	ttatcagttt	ctctgggtcc	43680
tttttagagg	taatcttttt	gtatgcaact	gtgcacatgt	atgtttacac	aaatggtagg	43740
gtattgtata	tatatacagt	tactgcagta	tgotattttg	actttgcttt	attgctttta	43800
ttttctgcct	gacatatgtt	aatgtggcca	aatagggcat	ttcccagaca	atccttatag	43860
tttcattctg	tgctacttta	aatgctgttc	tccttccctg	aaatgtcctt	tttttttttc	43920

ttttcttttt	ttttttttga	gacagagtct	cgctctgtca	cccaggctgg	agtgcagtgg	43980
cacaatctca	gctcactgca	acctccacct	ccccggttta	agccattccc	ctgcctcagc	44040
ttcccagagta	gttgggacta	caggcgctg	ccaccacgcc	cggatacttt	tttgtttggt	44100
tgttttagta	gagacaggg	ttcactgtgt	tagccaggat	cgcctcaatc	tcctgacctc	44160
ctgatctgcc	tgccctcagc	tcccaaagt	ctgggattac	agacgtgagc	cactgcgcct	44220
ggctgtcctt	tcttttaatc	ttttcaaatt	ctagccagtt	catagcccac	ttgctctgta	44280
aagctttttc	agtgatcttt	cttttgtacc	atgcagtctt	ttaaaaaatc	tacagtttta	44340
ttataagttg	ctaattatat	ttaataaata	cctaaatgg	agaaattgat	tcatttttta	44400
atctccgtag	tgccaggaac	tgtggctcat	gtctgtaatc	cctgcacttt	gggaggctaa	44460
ggcgggagga	tcgcttaagg	ccaggagctt	gagcccagcc	tgtggaacat	attgagacct	44520
cgtctctaca	aaaaataaat	tagctgggca	tagtgggtgc	tgtatgtagt	gtcaactact	44580
caggagactg	aggcagaagg	ataacttgag	ctcaggagg	tgaggccaca	gtgagccatg	44640
atggtgccac	tgtacttcag	cctgggtgac	agagcaagac	caagaccgat	aattatctcc	44700
atggtatttg	ttttggacgt	attagatatt	cagtaaatat	tttctggtga	tgatagtgat	44760
atctaactac	ctcaggttac	ctaacctgaa	taaggcctat	ttttaacata	gccattcact	44820
tgttaatttc	tgttattttc	gtgacttcat	ttttactaat	tatatattta	aagataattt	44880
tatataaatg	caaaaacttt	ttatcatgta	acaatctggt	aacaccttga	gtaatttgtc	44940
atttcacatg	cacattttat	ttgttcatta	tcttttttct	ttttgtttat	tttttctggt	45000
ttttaccttc	tctgttgcat	agagaagatg	caaccagttt	tgccattttc	caagtcacaa	45060
acggacgtct	ataatgacag	tactaacttg	gcatgcgcga	atggacatct	ccagtcaggt	45120
gggttttaggt	taactaacat	aaaataataa	agcttgcatg	caaagtaaga	gtcttactta	45180
acctgataat	gttgatttga	ttttataatg	gtaggtttaa	ttgttcatgt	tttcacaggg	45240
ctgtgttgaa	aatatagtac	ctttatcatc	atagtatata	aacatgcaac	aagtcaggta	45300
tagtcattct	aaacttgctt	tatttgcata	ttttgcaacc	tctgggtttt	cagattataa	45360
aatatgaagg	tatacattcc	cgtaagtgat	taaaataatc	tttttttctg	ttgtggcttt	45420
tgtgttttta	tatggatagc	tatactgatt	tttttcatga	aatagtgttt	tctaagacac	45480
attttaattt	tttatgtatc	aatttttatt	gcataatcaat	tttcatttat	agtatatgtg	45540
aatgtcttat	agtacttgat	agaagaaatt	tgaaacttca	catagtgagg	agaaacatta	45600

cagtattatt	tgctatggat	taactctatc	cttcataaac	ttttggatat	caacttctac	45660
agaagattaa	atgtctttct	gttttgtttt	ctttgggtag	gggactctcc	atttcttagg	45720
ttctctgaca	tctgagctag	ctttttgttt	aattaaaact	ttttttttag	ttgacagata	45780
atgtacctgt	tcatggggta	cacagtgatg	ttttgataca	tatagtatat	ggtgattaga	45840
tcaaggtaat	tatcatttcc	atcatcccaa	acatttatca	ttccttgtgt	tggaaacatt	45900
cagtattttc	cttctagcat	ttgaaatgat	ataatatatt	atattgttac	ctataatcat	45960
cctgcagtga	tgtagaacaa	tggagcttac	tctttctatc	tagctatacc	tttgtatcct	46020
ttccctaccc	ttttcagcct	ctagtatcct	ctgttctact	ttttatttct	atgagattaa	46080
ctttttaaaa	ataatattct	gtgtgtatat	ataccacatt	ttctttattc	attcgtctgt	46140
tcttggacac	ctaggttgat	tctgtatcct	ggctgttgtg	aatagtgtctg	cagtaaacat	46200
ggggtgtaga	tgtctgttcg	atatgatgat	tttctttcct	ttggataaat	tcccaatagt	46260
gggattggtg	ggtcatatgg	tagttctgtt	tgtagttttt	gaggaacctc	catactcttc	46320
tccatagtag	ctgtactagt	ttagattccc	aaaagtagcg	tataagagtt	cactttctac	46380
acatcctcac	cagcatttgg	tacttttttt	gtctttttga	taatagccat	cctaactggg	46440
gtaagatacc	ttgttgtggt	tttgatttgc	atttcocctca	taattaacga	tgttgagcat	46500
tttttcatat	ctttgttttt	gacggtccag	ctagttttat	tggttacttt	tttttttttt	46560
tttttagcaat	agtttcatgt	gcagggtcca	aattatatatt	aatattactt	ttttcaaatt	46620
ataccagaac	aaccttaggt	taggctataa	aacaactgcc	ttctttttct	tttggcataa	46680
taattatata	tatattttta	attacacttt	acgttctagg	gtgcatgtgc	acaacatgca	46740
ggtttgttac	atatgtatac	aagtgccatg	ttgggtgtgcc	gcaccatta	acttgtcatt	46800
tacattaggt	atatctccta	atgctacccc	tcccctctcc	cctcacccca	cgacaggccc	46860
catttgttga	tgttccccac	cctgtgtcca	agtgttctca	ttgttcagtt	cccatctgtg	46920
agtgagaaca	tacggtgttt	ggttttctgt	ccttgcgata	gtttgctcag	aatgatggtt	46980
tccagcttca	tccatgtccc	tacaaaggac	atgaactcat	cctttttgat	ggctgcttag	47040
tattccatgg	tgtatatgtg	ccatattttc	ttagtccagt	ctatcattga	tggacatttg	47100
cattgattcc	aagtctttgc	tattgtgaat	agtgccacag	taaacataca	tgtgcatgtg	47160
tctttatagc	agcatgattt	ataatccttt	gggtatatac	ccagtaatgg	gatggctggg	47220
tcaaattggta	tttctagttc	tagatccttg	aggaatcgcc	acactgtctt	ccacaatggg	47280
tgaactagtt	tacgttccca	tcaacagtgt	aaaagtgttc	ctgtttctcc	acatcctctc	47340

cagcacctgt	tgtttctga	gtttttaatg	atcgccattc	taactgggtg	gagacgggat	47400
ctcattgtgg	ttttgatttg	catttctctg	atggccagtg	atgatgagca	ttttttcatg	47460
tgtcttttgg	ctgcataaat	gtcttctttt	gagaagcatc	tgttcatata	ctttgcccac	47520
tttttgatgg	ggttgtttga	tttttttctt	gtaaattttt	ttaagttctt	tgtagattct	47580
ggatatcagc	cctttgtcag	atgggtagat	tgcaaaaatt	ttctcccat	ctataggttg	47640
cctgttcaact	ctgatggtag	tttcttttgc	tgtgcagaaa	ctcttttagt	taattagatc	47700
ccatttgcca	attttggcct	ttgttgccat	tgcttttggt	gttttaggca	tgaagtcctt	47760
gcccattgcca	gtgcccattg	cagtgtcctg	aatgggtattg	cctagatttt	cttctagggt	47820
ttttatagtt	ttagaactaa	catttaagtc	tttaatccat	cttgaattaa	tttttgtata	47880
aggtgtaagg	aagggatcca	gtttcagctt	tgtacgtttg	gctagccagt	tttcccagca	47940
ccatttatta	aataggaaat	cctttcccca	tttattgttt	ttgtcagggt	tgtcaaagat	48000
cacatgggtg	tagatgtgtg	gtattatttc	tgagggctct	gttctgttcc	attggtctat	48060
atctctgttt	tggtatcagt	accatgctgt	tttgattact	gtaccttcgt	agtatagttt	48120
gaagtcagg	agcacgatgc	ctccagcttt	gttcttttgg	cttaggattg	tcttggcaat	48180
gcaggctctt	ttttggttcc	atatgaactt	ttaaagtagt	ttttccaatt	ctgtgaagaa	48240
agtcatttgt	agcttgatgg	ggatggcatt	gaatctataa	attaccttgg	gcagtgtggc	48300
cattttcatg	atattgattc	ttcctatcca	taagcatgga	atgttcttcc	atttgtctgt	48360
gtcctctttt	atttcattga	gcagtgggtt	gtagtctctc	ttgaagaggt	ccttcacatc	48420
ccttgtaagt	tggattccta	ggtattttat	tctctttgaa	gcaattgtga	atgggagttc	48480
actcatgatt	tggctgtttg	tctgttattg	gtgtatagga	atgcttgtga	tttttgcaca	48540
ttgatattgt	atcctgagac	tttgctgaag	ttgcttatca	gcttaaggag	attttgggct	48600
gagatgatgg	ggttttctaa	atatacaatc	atgtcatctc	caaacagggg	caatttgact	48660
ttcttttctt	aattggatac	cctttatttc	tttctcctgc	ctgattgccc	tggccagaac	48720
ttccaacact	atgttgaaca	ggagtgggtg	gagagggcat	ccctgtcttg	tgctagcttt	48780
caaaggggaat	gcttccagtt	tttgcccatt	cagtatgata	ttggctgtgg	gtttgtcata	48840
aatagctctt	attattttga	gatacatccc	atcaatacct	agtttattga	gagtttttag	48900
catgaagggc	tgttgaattt	tgtcaaaggc	cttttctgca	tctattgaga	taatcatgtg	48960
gtttttgtct	ttggttctgt	ttatatgatg	gattacgttt	attgatttgc	atatgttgaa	49020

ccagccttgc	atcccagggga	tgaagccaac	ttgattgcgt	tggataagct	ttttgatgtg	49080
ctgctggatt	cggtttgcca	gtatttttatt	gaggattttt	gcgttgatgt	tcatcagggga	49140
tattgggtcta	aaattctctt	tgttgtgtct	ctgccaggct	ttggatatcag	gatgatgctg	49200
gcctcataaaa	atgagttagg	gaggattccc	tcttttttcta	ttgattggaa	tcatttcaga	49260
aggaatggta	ccagctcctc	tttgtacctc	tggtagaatt	cagctgtgaa	tccatctggt	49320
cctggacttt	ttttggttag	taggctatta	attattgcct	caatttcaga	gcctgttatt	49380
ggtctattca	gggattcagc	ttcttcctgg	tttagccttg	ggaggctgta	tgtgtccagg	49440
gatttatcca	tttcttctag	attttctagt	ttatttgagt	agagggtgtt	atagtattct	49500
ctgatggtag	tttgcatttc	tgtggaatcg	gtggtgatat	cccccttatc	atttcttatt	49560
gcatctattt	gattcttctc	tcttttcttc	tttattagtc	ttggtagcag	tctatcagtt	49620
ttgttgatct	tttcaaaaaa	ccagctcctg	gattcattga	ttttttgaaa	ggttttttgt	49680
gtctctatct	ccttcatttc	tgctctgata	ttagtatttt	cttgccctct	gctagctttt	49740
gaagggtgtt	gctcttgctt	ctctagttct	tttaatgggtg	atgttagggg	gtcaatttta	49800
gatctttcct	gttttctctt	gtgggcattt	agtgcataaa	atttccctct	acacactact	49860
ttaaatgtgt	cccaaagatt	ctgatatgtt	gtgtctttgt	tctcgttggg	ttcaaagaac	49920
atctttattt	ctgccttcat	tttcttatat	accagtagt	cattcaggag	cagggtgttc	49980
agtttccatg	tagttgtgtg	gttttgctgt	agtttcttaa	tcctgagttc	tagtttgatt	50040
gcactgtggc	ctgagagaca	gtttgttgta	atttctgttc	ttttacattt	gctgaggagt	50100
gcttttagttc	caactatgtg	gtcaattttg	gaataggtgt	gggtgtgggtg	tgagaagaat	50160
gtatattctg	ttgatttggtg	gttttagagtt	ctgtagatgt	ctattagggtc	cacttggtgc	50220
agagctgagt	tcagttcctg	gatctgtctt	gttgatctgt	ctaataattga	cagtgggggtg	50280
ttgaagtctc	ccagtattat	tgtgtgggag	tctaagctct	tttgtaggtc	tctagggact	50340
tgctttatga	atctgggtgc	tcctgtattg	ggtgcatata	tatttaggat	agttagctct	50400
tcttgttgaa	ttgatccctt	tagcattata	tgatggcctt	ctttgtctct	tttgatcttt	50460
gttggtttaa	agtctgtttt	atcagagagt	tggattgcaa	accctgcttt	ttttgttttc	50520
catttgcttg	gtagatcttc	ctccatccct	ttattttgag	cctatgtgtg	tctctgcacg	50580
tgagatgggt	ttcctgaata	cagcacactg	atgggtcttg	actcgttatc	caatttgcca	50640
gtctgtgtct	tttaattgga	gcatttagcc	catttccatt	taaggttaat	attgttatgt	50700
ttgaatttga	tcctgtcatt	atgatgttag	ctggttattt	tgctcgtag	ttgatgcagt	50760

tctcttcctag	cctcgatggt	ctttacaatt	tggeatgttt	ttgcagtggc	ttgtaccggt	50820
tgttcatttc	catgttcagt	gcttccttca	ggagctcctg	taagcaggcc	tggtagttag	50880
aaaatctgtc	agcatttgct	tgtctttaaa	ggattttatt	tctccttcac	ttatgaagct	50940
tagtttggtc	ggatatgaaa	ctctgggttg	aaaattcttt	cctttaagaa	tgttgaatat	51000
tggcccccac	tctcttctgg	tttttagagt	ttctgccaa	agatcagctg	ttagtctgat	51060
gggcttccct	ttgtgggtaa	cccgaccttt	ctctctggct	gcccttaaca	ttttttccta	51120
catttcaact	ttggtgaatc	tgacaattat	gtgtcttgga	gttgctcttc	ttgagtagta	51180
tctttgtggc	attctctgta	tttctgaat	ttgaatgttg	gcctgccttg	ctaggttggg	51240
gaacttctcc	tggataatat	cctgcagagt	gttttccaac	ttggttccat	tctccccgtc	51300
actttcaggt	acaccgctca	gacatagatt	tggtcttttc	acatagtcct	atatttcttg	51360
gaggttctgt	tcgtttctct	ttactctttt	ttctctaaac	ttctcttctg	gcttcatttc	51420
attcatttga	tcttcaatca	ctgataccct	ttcttccact	tgatcgaatc	ggctactgaa	51480
gctcatgcat	gcatcacgta	gttttctgtc	catggttttc	agctccatca	ggccatttaa	51540
ggtcttctcc	atgctgttta	ttctagttag	ccatttgtct	aatctttttt	caagggtttt	51600
agcttctttg	caatggtttc	gaacatcctc	ctttagctcg	gagaactttg	ttattaccca	51660
tctgtctgagg	cctacttctg	tcagcttgtc	aaagtctttc	tctgtctagc	tttgttccgt	51720
tgtctgtgag	gagctgtgtt	cctttggagg	agaagaggcg	ctctgaattt	tagaattttc	51780
agcttttctg	ctctggtttc	tcccatctt	tgtggcttta	tctacctttg	gtctttgatg	51840
atggtgacgt	acagatgggg	ttttgggtgt	gatgtccttt	ctgtttgtta	attttccctc	51900
taacactcag	gaccctcagc	tgcagggtctg	ttggagtttg	ctggagggtcc	actccagacc	51960
ctgtttgcct	gggtatcacc	agcggagggt	gcagaacagc	aaatgttgca	gaacgacaaa	52020
tgttgctgtc	tgatccttcc	tctggaacct	tctgtctgaca	ggggtaccca	ggtatatgag	52080
gcgtcagtc	gcccgtatgg	ggagggtgtct	cccagttagg	ctacttgggg	atcagggacc	52140
cacttgagga	ggcagctctgt	ccgttcgccg	atctcaaact	ccatgctggg	agaactacta	52200
ctctcttttag	agctgtcaga	cagggacctt	taagtctgta	gaagttactg	ctgccttttg	52260
ttcagctatc	ccatgtcccc	agagggtggag	tctacagagg	cagtcaggcc	tccttgagct	52320
gtggtggact	ccaccagtt	caagcttcct	agctgctttg	tttaccact	caagcctcag	52380
caatggcaga	tgccccgcc	tccagcctct	ttgccgcctt	gcagttcgat	ctcagactgc	52440

tgtgctagca	gtgagcgagg	ttccgttggc	atgggaccct	ctgagccagg	catgggatat	52500
aatctcctgg	tgtgccgttt	gctaagacca	ttggaaaagc	ccagtattag	ggtagggagtc	52560
tccctatttt	ccaggtacca	tatgtcacgg	cttcccttgg	ctaggagagg	gaattcccca	52620
acccttgcg	cttcccaggt	gaggcaatac	cccgcccttc	agctcacact	atgtgggctg	52680
caccactgt	ctgacaatcc	ccagtgagat	gaaccagtt	cctcagttgg	aaatgcagaa	52740
atcagctgtc	ttctgtgtcg	ctcacactgg	gagctgcaga	ctggagctgt	tcctattttg	52800
ccatcttgg	accctgcctt	cttcattcat	atgtaataca	aaacttctaa	ggtttttagtg	52860
gagaagagat	agagtaaaag	gatttctata	gacagaagaa	acagttggtc	atcaactctt	52920
tcccttgtgg	cttcacatct	ccctaagga	cttcttatgt	tggtttggtc	ttacagtata	52980
gtcagtggca	gtttccctct	tcccattcct	tcctctgtct	gatttaaaat	gctgtttcaa	53040
gtatcgatat	aagtattttg	cctgtttctt	ttagcgtggc	tgtgaagggc	tgacattttc	53100
agaaggcact	tactgaaaaa	aaaaaaacaa	agaaatgtaa	gagtccatca	catataaata	53160
gttaagtttc	taaaatatgt	atttgagatc	ccagtaattc	tactaggata	aatagcaaaa	53220
attctccagc	cctgaagagt	tggtctgtct	ttcctttcct	tgttatcttg	attctctttt	53280
atttcatttc	acatgctaga	ccccatcatg	ttttcctgct	gcattcccca	ctccaccctc	53340
caaagcaatg	ttctcttcct	gcaacctgtt	tgaaaaaatt	gaactttgtc	ttatttcata	53400
tccctagtag	aagttgaata	tcctaatat	gaaatctgaa	atccaaaatg	ctctaaaatt	53460
cgactctttt	ttttttgaga	tgaagtcttc	ctctgtcact	caggctggag	tgcgggtggcg	53520
caatcttggc	tactgcaac	ctccgcagcc	tcccctaacc	ccatccctgt	gggttcaagc	53580
gattctccca	cttcactctc	ccaagtagct	gggactacag	gcacctgcca	ccacacccaa	53640
ctaatttttt	gtatttttag	tagagatgag	gtttcaccat	gttggccaag	ctagtctgag	53700
actcctgacc	ttaagtgatc	cacctgccct	ggcctcccaa	agtgttgga	ttacagtcac	53760
gagccactgt	gcctagccaa	atatccaaaa	ctttttgagc	gctgacatga	tgctcaaagg	53820
aaatactgga	gtattttgca	ttttggattt	ttgggttaag	gatgctgaac	cagtaagtat	53880
aatgtaaaaca	ttgcaaaatg	caaaaaaatg	taaaaaccct	aagcagttct	gggtcccaagc	53940
atttcgcata	agggatactc	aacccataat	cttttctttt	ctgttttctg	gttggaaggg	54000
catattggct	ttatagctaa	tttacacagc	attgatgttt	aatacacagt	gagtccaagt	54060
agatcactca	gacctattag	tagtttatta	gtgtcactca	cttctggaac	attctgtgat	54120
gttttacttg	gatgagttct	ttcacctctc	ttgggaaata	gtcataccaa	agtctgctta	54180

ttactacaat	gtgttatcta	ttatacattg	tccaattttc	tactcaaaat	tactagacag	54240
gcaaagaaaa	agtaaagtat	aacccttact	caagaaaaaa	aagcaatcag	tagaaactgt	54300
gagtgggccc	agatattgga	tttagcagac	aaagacttca	aagcagctat	tataatatgt	54360
ttaaagaatt	gaaagaaaat	atggtatcag	ttaaacagga	aatctaagta	gatgatataa	54420
actagacaat	aaaaataata	ttctaaagtt	gaaaagtgtg	gttactgaaa	ttaaaaattt	54480
acagacaaca	gcctcaacag	cagattagag	atagcaaaaag	aaagattcag	tgactttgaa	54540
tgacaggtctg	tagaaattac	taaattgatg	actctcacgt	agcaactttc	acccgtagtt	54600
tggtttcata	tacaatgctt	taactctttt	tgctattttt	tctcttactt	tctgtgatgg	54660
aaacatttttg	tgcttttgatt	ttaatgggtg	tatcacagtt	atatacatct	ttacaaactc	54720
atcaaattgt	atactttatt	tatatattta	tttttttttg	agatagggtc	tgactctttc	54780
tcccaggctg	gagtacagct	gaaattagtt	gggactacag	gtgggcacca	ctgtgccagc	54840
taatttttgt	attttttgta	gagatggggt	ttccccatgt	tggccagtct	ggctctcaaac	54900
tcctgggctc	aatcatttct	gccgcctcag	actcccagag	tgttgggatt	ataggtgtga	54960
gctactgtgc	ctggctggta	tagctttttt	tttttttttt	tttttttttg	tgacagagtc	55020
tcactctgtc	gcccagggtg	gagtgcagtg	gcgcgatctg	ggctcactgc	aacctctgcc	55080
tcccgggttc	aaccatttct	cctgcctcag	cctcctgagt	agctgggact	acagatgcgt	55140
gccaccatac	ccagctgatt	tttgtattct	tagtagaaac	agagtttcac	catgttggcc	55200
aggatgatct	cgatctcctg	acctcgatcc	acctgccttg	gcctcccaca	gtgctggcat	55260
tacaggcctg	agccaccatg	cctggctggt	ggtatacttt	aaatggatgt	aattcattgc	55320
agattatacc	tcagtaaatt	tttatttagt	ttttgagact	gagttgctct	gttgcccagg	55380
ctggagtgca	gtggcacgat	ttcagctcac	tgcaacctct	gcctcctggg	tttaagcgat	55440
tttcctgtct	cagcctcccc	ggtagctgga	attagagggtg	tgtgccacca	tgatcagatt	55500
atttttgtgt	tttttagtaaa	gatgggggtt	caccatgttg	gccaggctgg	tctcgaactc	55560
ctgacttcag	gtgatccacc	cgctcgggcc	tctcaaagtg	ctgggattac	aggtgtgagc	55620
caccacgcct	ggccttaatt	tttaaatact	gtaaggctta	taaagaaaag	aatattcccc	55680
ttctgtttct	ttcctctcac	gtagcaacct	tcacccttag	tttggtttca	tatacagtgt	55740
tttaactctt	tttgctattt	ttttctctta	ctttctacta	tatttccaaa	tacaatgctt	55800
ctataatgat	tccttttttt	tttctatcag	tttttgataa	tcattgactc	cttatgggtca	55860

aagaagactt	aattcccttc	tgtcactctt	catatattaa	tataactaat	atatatatg	55920
tttagttttg	atataactaa	aaataactat	atgtatatat	aactatatat	gtataagcta	55980
tctgtatata	gttatatatg	tatatgcaca	tacatatata	gttacatatg	tatacacaca	56040
tacatatata	gttacatatg	tatatgcata	tacatatata	gttacatatc	tatatgcata	56100
tacatatata	gttacatatg	tatatgcata	cacatatata	gttacatatg	tatatgcata	56160
tacatatata	gttacatatg	tatatgtata	tgtatataca	tatatagtta	tatatgtatt	56220
agttatatca	aaataactat	gtatatatag	ctgtaaagt	atatataaac	tatatgtata	56280
cagttatata	tgtatttgta	tatgtgtgta	tacatatata	gttttttttg	tttttttttt	56340
ttgttttttt	tgtttttttt	tggagatgga	gtcttgcct	gtccccagg	ctggaatgca	56400
gtggtgctat	cttggcttac	cgcaacctct	gcttcccagg	ttcaagcaat	tctcctgctt	56460
cagtctcccg	agtacctggg	attacaggca	cgtggcacca	cgccaggcta	attttttgta	56520
tcttttagtag	aaatgggggt	tcaccatggt	ggccaggctg	ttctcaaact	cctgacctcg	56580
tgatccaccc	gcctcggcct	cccaaagtgc	tgggattaca	ggcgtgagca	ccgcgcctgg	56640
tccttataca	tatatagtta	tatataccta	tagttatttt	tagttatatc	aaaataacta	56700
tatgtatatg	taactatata	tatagtatat	atagtatata	tatactatat	agtgtgtata	56760
tatagtatat	atatactata	tagtgtgtat	atatagtata	tatatagtgt	atatatcgta	56820
tatacactat	atactatata	gtgtatatat	agtatatgta	gtatatatag	tatatatagg	56880
atagtatata	tagtatatat	agtgtatata	tactgtatat	ataggtgtac	atagtatact	56940
atatagtata	catatagtag	actgtatagt	atatagtata	tagtatatat	agtatacata	57000
gtatactata	tatagtagag	tatacatagt	atactatata	gtatatagag	tatatatata	57060
gtatactata	tagtatatag	agtatatata	cagtatacta	tatcgtgtgt	atagagtata	57120
tatacagtat	actatatacg	gtgtatagag	tatatatata	gtatactata	tcgtgtgtat	57180
agagtatata	tacagtatac	tatatcgtgt	gtatagagta	tatatacagt	atactatatc	57240
gtgtgtatag	agtatatata	cagtatacta	tatcgtgtgt	atagagtata	tatacagtat	57300
actatatacg	gtgtatagag	tatatatata	gtatactata	tagtgtgtat	agagtatata	57360
tacagtatac	tatatagtgt	gtatagagta	tatatacagt	atactatata	gtgtgtatag	57420
agtatatata	cagtatacta	tatagtgtgt	atagcgata	tatacagtat	actatatagt	57480
gtgtatagcg	tatatatata	gtatactata	tagtgtgtat	agcgatatata	tacagtatac	57540
tgtatagtgt	gtatagcgta	tatatacagt	atactgtata	gtgtgtatag	cgtatatata	57600

gtatactgta	tagcgtatag	agtatatata	tagtatactg	tatagtgtat	agagtatata	57660
tatagtatac	tgtatagtat	atagagtata	tatagttaca	tatacatata	gttatttttta	57720
gtaatatcaa	aaaaccatag	ctaagattttt	tatgatttag	caaaatattg	tttacccecaa	57780
agccacgtat	tatatattaa	tgatatTTTT	tcccttttga	gtttttcttc	cccacagggt	57840
ttctgatctt	tttttttggt	ctctaacttt	atattatagg	tatatactct	tttctctggt	57900
taatttgctt	tttaactaga	agacctcctt	tagtagttct	tttagtgaat	atctgtgggt	57960
aaactcttag	tctttccagt	gtctgaaata	ccctcatTTT	aatgtgacag	gtcacttttc	58020
tcacctcaac	atTTTggaca	tacctcatTT	ccaacaaaca	tttgttactt	ctgatagaga	58080
gtctgctgtc	attctcatac	ttgttccttg	tggaaatgtc	tatatataaa	gtctcagaac	58140
tatatatTTT	taatatatat	acatatTTTT	atttatTTaa	aaatgtataa	ttaagttagt	58200
agtagtagta	gtagtagtag	tagtagtagt	agtagtagtt	ttttttTTTT	tagatggagc	58260
ctccatctcc	cgggttcaag	tgattctcct	gcctcagcct	cctgagtagc	tgggattaca	58320
ggcacacacc	accatgcctg	gctaattTTTT	gtatTTTgag	tggagatggg	gtttcaccat	58380
gttgGCCagg	ctgatcttaa	actcctgacc	tcagatggac	cacctgcttt	ggcctcccaa	58440
agcgctagga	ttacaggcat	gaaccactgc	gtctggcctt	aagttattat	tgactattga	58500
tgccctgTTg	ctctatcaag	tagtaggtct	tatttattct	agaacctcgt	tttaagggtat	58560
cctgaatgtg	gctTTTTTTTT	ttttttttcc	taagacggag	tctcactctg	ttgccagggc	58620
tggagtgcaa	tggagtgatc	ttggctcacc	gcaatctctg	cctccttggt	tcaagcaatt	58680
ctcctgcctc	agcctcctga	gtagctgaga	ttacaggcgc	acgccaccac	acctggctaa	58740
tttttgTgtt	tttagtagag	acgggggtttc	accacattgg	tcaggcttgt	ctcaaactcc	58800
tgaccttgTg	atccgcccac	cttggcctcc	gaaagtTctg	ggattacagg	cgtaagccac	58860
cacgcccagc	ctaattgtgga	ttttgtTTTT	ttttttaact	tgaattttat	ttataacttcc	58920
tgatTTTata	atagttaact	tgtagtaaaa	gtaaactggt	tgtctaaata	aataagctga	58980
tttcctatac	tagtttagag	tcttttagttt	tcttaaaccg	tattaagaga	tttgattaca	59040
aaagtaaaat	tgaggctttt	gtctTTTTTT	ttggccactc	taatttgctg	actgacttaa	59100
catttgTctt	cacagtaaac	attgtaaaata	aacactgatg	tttgttgcgt	agtatgtaaa	59160
aatattatTT	aggagtggTg	agatttaaac	atttacaaac	ctgtaatata	tatagtacag	59220
caacactgga	caaataacct	tttcaactta	atcagaggTg	ttctataaat	gttcatttgc	59280

atatatacag	cccatataat	acttagcata	tgtaaaagca	catctttata	ctctttcaat	59340
taaaattgat	tattattatt	atTTTTTTTT	tgagacagag	tccctctcag	tcacccaggc	59400
tggattgcag	tggtgtgac	agggtcact	gcaatcgggt	agctggaact	acagggatgt	59460
gccaccacgc	ccagctatTT	ttgtatTTTT	agtggagatg	gggtttcacc	atgttgGCCA	59520
ggctggctct	gaactccata	cctcaggtga	tctgcccgcc	tcagcctccc	aaatagctgg	59580
gataacaggc	gtgagctact	gcacccggcc	taaaattgat	tagtttttaa	gaccctcttg	59640
gtgggaacag	actttcatag	tgtttatagt	tataatagtt	tggaatcaaa	gtagtcgagt	59700
gctactTTTT	tttctctgt	acatcctatc	cctaattgct	ttactatatt	ttgtttctta	59760
tttgagagct	ttatTTgctt	cataactTTT	TTTTTTTTTT	TTTTTTTTtg	agacagagtt	59820
tcactctgtc	accagggctg	gagtgcagta	gtgcaatctc	ggcttactgc	aacctccgcc	59880
tcctgagttc	aagtgattct	cctgcctcag	cctccgaagt	agctgggatt	ataggccccc	59940
accaccacgt	ctgactgatt	ttttgtatTT	ttagcagaga	cgaggtttca	ccatgttggc	60000
cagcctggtc	ttgaactact	gacatcaggt	gatccgtcca	cctcgatctc	ccaaagtgt	60060
gggattacag	gtgtgagcca	tcgcgcccgg	cgcttcataa	cctTTTTata	ttgtctttct	60120
gttcccaaaa	atgaagtagt	ttcttcttaa	aaatactgtt	ttttgttttt	ttgttttttt	60180
ttttatttga	gatggaatct	ctctctgtcg	ccaggctgga	gtgcagtggc	acgatctcgg	60240
ctcacttcaa	cctctgactc	cgtggttcaa	gcaattctcc	tgcctcagcc	tcccaggtag	60300
ctgggattac	aggcacaat	actgttttta	ataagaaaat	tagaataggg	acaggcacia	60360
tggctcactc	ttgtaatccc	agcactttgg	gaggccaagg	tgggaggatt	gcctgagccc	60420
aggagttcga	gaccagccta	ggcaacatag	tgagtgagac	cctgtttcta	taaaaaaaaat	60480
ttttaattag	tccggtctgg	tggtgcacaa	ctgatgtccc	agctacttgg	gagtttgagg	60540
tgggaggatg	gatcacttga	gccaggaag	attaaggcca	caatgagctg	tgatcatgcc	60600
actgcactcc	agcccaggcc	acaggtttca	aaaaaagaga	aaattagaat	tatacgttag	60660
ggggaaaaaa	cctaataacc	atattcctat	ctacctagtg	accacccta	tgaagatcct	60720
ggtacatggt	ctcattgaaa	tcttatTTtg	aaatatTTTT	gaattaaaaa	aatatattat	60780
taccttggtt	ttacaaatgt	ttgcttgtct	ttatgttcag	ctacaatttt	ctaatcacia	60840
tggttttact	ttttccttgt	cagaaaagtg	agctgttcca	aaaagaaaag	acccttaac	60900
acacactagt	aattcactgc	ctcgttcaaa	aacagttatg	aaaactggat	ctgcaggcct	60960
ttcaggccac	catagagcac	ctagttacag	tggtttatcc	atggtttctg	gagtgaacaa	61020

gggatctggg	cctgctccta	ccactcataa	ggatattctgg	gacagtaact	ttaattgctg	61080
tcttttttgca	aatagaaaaa	tttttaagat	actattcctg	cttaagttga	tcataagtac	61140
tttataatac	tttagagaat	ggataagttt	ccataaagtt	aaatTTTTTT	TTTTTTTTTT	61200
tgagactctt	gttttccagg	ctggagtgca	atggcgcgat	ctcggetcat	tgcaacctcc	61260
gcctcccagg	ttcaagcgat	tctcctgcct	cagcctccca	agtagctggg	attacaggca	61320
tgcaaccacca	cgtcgagcta	attttgtatt	tttagtagag	acggggtttc	tccatgttgg	61380
tcaggctggg	ctcaaactcc	cgacctcagg	tgatccgcct	gcctcagcct	cccaaagtac	61440
tgggattaca	ggcgtgagcc	accgtgcccc	aagttaattt	TTTTTTaaat	ccacagggca	61500
acttgctata	aaactTTTTT	tggacctcat	tagacattta	ataccaagtt	ttgcttacct	61560
agtctgtttt	atataatgct	gtatatTTTca	TTTTaaattt	atTTTTaaaa	ttatttcaat	61620
ctttgtaaat	gttttatacc	agtaacatac	ctagagtttt	gtcatgattc	taacaagggg	61680
taaaatttgt	ttttatTTTg	taaaaactaa	atctctggat	aaaatcctct	ataactgact	61740
taatgccaaa	tatagtTctc	tagtgaatac	agTTTTacct	tcaggtaaat	aaatatacaa	61800
tttatggatt	tttataccct	ttttcctatt	tttaaagctt	gaattctgtg	aactttaagg	61860
ttaacttatt	tatgaaaagt	gtaaatgtta	ggttgtattt	tcatattaaa	atTTTgtatc	61920
ctttaagggg	actccgaaaa	caaataggac	aaataaacct	tctacccta	caactgttac	61980
tcgtaagaaa	aaagacttga	agaatttttag	gaatgtggac	agcaaccttg	ctaaccttat	62040
aatgaatgaa	attgtggaca	agtaagtttt	gccatctaaa	tgTTTTattt	tatagtTTTT	62100
atattttaat	tttacttata	aaacatgtca	ggagtgaat	agataataaa	taccttgtct	62160
ggtttacata	cttaattTTTT	atgatagttt	tcaattataa	atgtagaaaa	ccattgcttt	62220
actgatttaa	ctatagttta	acagttaaca	ttaaaaataa	ctatatgtca	tagggcttag	62280
gcttcactct	gtaataactg	ggccctgttt	gtatcgtaga	actaactgag	gtcttgTTTT	62340
ttagtggaac	agctgttaaa	tttgatgata	tagctgggtca	agacttggca	aaacaagcat	62400
tgcaagaaat	tgttattctt	ccttctctga	ggcctgaggt	aagaacttta	tattatcatt	62460
tttctataat	accatctgtt	actgaatcca	tagtagtagt	agtagtaaaag	aaatatTTTga	62520
gctataactaa	aataattaat	tcatataagg	taacaataga	tttaatgttt	taaaaaagat	62580
aaaacattaa	caattgatgt	tgagaagtta	ttatagaata	ggaaaatgga	tatgagttcc	62640
taatttccaa	tccttttata	tttttaggaa	aagtgggtag	tatgatataa	ttttgtTTTg	62700

ttttggagac	ggagtctcac	tcattgtgtc	gcccaggctg	gagtgcagtg	gcatgatctt	62760
agctcactgc	aacctctgcg	tcccagaatt	caagtgattc	tcctgcctca	gcctcccaag	62820
tagctgggat	tacaggcoatg	tgccactatg	cccagctaat	ttttgtaatt	ttagtagaga	62880
tggggttttg	ctatatgggc	caggctggtc	tttcatacct	gatgtcaagt	tatccaccca	62940
tttcggcctc	ccagagtgtt	aggattacaa	gtgtgagtca	ctacatctgg	ccaaattttg	63000
atatcaagg	gagagagatt	taaaattaag	ataaggtaga	aaaattagcc	tagtgtgggg	63060
gcgcacgctt	gtagtcccaa	ctactgggga	ggctgtggca	ggagaattgc	ttgaaccagg	63120
aggcagaggt	tgcagtgagc	caagatggca	ccactgcact	ccatcctggg	tgacagagcg	63180
agatgtcatc	tcaaaaacaa	aacaggccgg	gtacggtagt	tcacgcctgt	aatcccagca	63240
ctttgggagg	ctgaggcggg	cagatcacga	ggtcaagaga	ttgagaccat	cctggctaata	63300
atggtgaaac	cctgtcttta	ctaaaattac	aaaaattagc	tgggcttggt	ggtgtgtaac	63360
cccagctact	cgggatgctg	aggcaggaga	atcgcttgaa	cccgggaggc	ggaggttgca	63420
gtgagccgag	atcacgccac	tgcactccag	cctggcgaca	gagcgagact	ccgtctcaaa	63480
aaaataaata	aaaatttaaa	aagataaata	cataaaaata	aataaataat	attaagagaa	63540
ggaaatcagg	caggtagtgg	cccctgacac	aatgagtttt	cccagaattg	gattgcttgg	63600
aaatgccgct	caaagagtgt	ggtaaactcc	atcgaaggct	aaataccaac	gtgacagtga	63660
taataaacia	gtactttatg	ggaaagtttt	ttttttaatt	attttttaaaa	agagagaaat	63720
tgtactggag	aaaagaggaa	ttcaggtaga	ataattcttt	tttttttttt	ttttgagatg	63780
gaattttgct	cctgttacct	aggctggagt	gcagtggctt	gatgttggct	cactgcaacc	63840
tctgcctttg	gggttcaagt	gattctcatg	cctcatcctc	ctgagtagct	gggtctatag	63900
gcatgcaccc	caacacctgg	ctaatttttg	tgtttttagt	agagataggg	ttttaccatt	63960
ttggccaggc	tggtctcaaa	ctcctggctt	cagatgatcc	gcccaccttg	gcttcccaaa	64020
gtgctgggat	tacaggcaga	ggccactgtg	cctggcatga	agaacaattt	ttaaaagaat	64080
gacttttaag	gatattaagt	catcaaagta	gatagagcca	ttaaatgatg	ggtagaaacc	64140
taatcttcca	tcccatattt	tatgacttat	taaggaagat	aggcaatctt	gcagttgata	64200
ataaatattt	ggctttccat	acttagcacc	gttttgaatt	tttcagttc	acagatggta	64260
tatttagtgc	tgttccaata	atattgcaaa	aaataaatct	gaagactcac	ttctgggtca	64320
cagttatttc	actattaaca	ttaaaatctt	acggacctac	ctgcaacatg	tagtggaat	64380
aagttgtgtg	gcacgttgtg	gtgggtgcat	tattaaataa	atgtgcaaag	gttttatggc	64440

tgtctcaatt	tttccatagt	cttgagtcac	tcacatactg	ccttttatggt	gtttatcact	64500
tgtaagtgag	taccaactgt	actagaactt	actgaatatt	tttcaagtct	attttactta	64560
ttactcattc	taagtacata	aaatcactgg	tctgatatgc	aagttagggt	ttcgctaata	64620
cccagtaaaa	taattagata	attattaaaa	tagaaagcat	ttgtcattat	gtgaactgga	64680
tttatgtcat	gtaccacctc	aggtcacact	ttgggaaacg	tagcttacgt	attgagctaa	64740
cattttgctg	gagttatata	agattagtaa	atgctataga	ctaaatattg	aaaaaaaagc	64800
ttgctttaaag	aggattttaca	taattgtaga	cctgaaaagg	atcttagcag	tcttctgata	64860
gattcattga	taaagaaact	agtcctaaga	gctaaaatga	cttgctcaag	attttagtta	64920
agtggtagac	ctaggaaata	cctttctaag	attagagctg	cctgggttagt	gctatagctg	64980
ccacctaat	cttacattta	tttattaagg	gacatgcttt	ttacggtaga	aaaatatata	65040
taaaattgtg	aaggtaatac	caaggaaata	tttaggctat	gcaagaaaag	tcattcagct	65100
tttgctttac	ccaggcctat	agagaacata	tctaaaataa	taagtaacat	gtgaatttga	65160
atntagcat	attttcatgg	tagtgacata	atgtgcaatt	aagaagctat	aataatgtaa	65220
aaatgccagt	tgcaagttat	ttcaaataaa	gttgaataac	ctgacctcag	gaggggcaga	65280
aaccttagta	aacttagcaa	actgagacct	aaggaagagc	aagatctgtg	gtagagccag	65340
gtctaaaggt	ttatatcctg	gtacaaaact	gtctatacat	caccacacct	gccaaatgga	65400
actttttttc	cttgtagttc	agatttccaa	gaaagggaat	attattgtcc	tagctcagcg	65460
tttcacaata	gactcacaaa	atgaatcttg	ttaagcctat	gaatgggtaa	gacttatccc	65520
ataacacgcc	tctattgtgt	aaaaaaatca	gctttattcc	taattccctg	tagtaggaag	65580
cttgcataga	ccacaataga	gacattgtta	gattatatca	agaaggtgag	tggcaggcat	65640
tcttgagtat	actggaattt	ttatttcatt	gtagaaacac	agaatcataa	ttagtcatct	65700
aaaactgggt	ttaaggagat	aatactaaag	atactcactt	gaggtagatt	gttctcagaa	65760
ttacttaaaa	taagtgcctt	aaatcttgct	cattcatgaa	gccatagccc	tggaaggaga	65820
tattgaaaaa	cttctccatc	ttccttcaat	cccataggac	catgttttta	attgtagttt	65880
aacattaaac	ataatttttg	aaagtagcca	atattaatct	ttgacaaaaa	actcaaacta	65940
aatagcagaa	aatagaacat	gccctctcct	ctcgcccaaa	agtaagacta	catttttaaag	66000
tatatcacag	tgatctgtct	aaatattttt	ctcaaataat	ttactaagaa	agcgtatatt	66060
cttctgtcct	tatttcatag	atctcttctg	tatcctgtgc	agcttgtcta	tcatttctgag	66120

cttttatatt	tcttattttg	ttctcttgct	gcttaaagt	gattctttga	gatttataga	66180
agaatgtgat	tgttctccag	ttttaaaaag	ctcttctcta	tctactaaaa	tttccatgaa	66240
tttctttctt	tctttttttt	tttttgattg	agacagtctc	gctgtgttct	ccagcgtggc	66300
cttaggctta	agtgatcctc	ctgtctcagc	ctcccaagta	gctgagatta	cagaaatttc	66360
catgaatttc	caaatatgat	gttatctttc	atataagctt	aaacataagc	ctttcttctt	66420
tcaaaaatgt	tagaacttac	ctctttaaat	aatcttataa	agccatttct	taatttttgt	66480
taccttttgc	aatctattaa	cttgcatatc	atagtaaagt	aacccttagc	tccatcatac	66540
cacacacctt	ctactcctgt	cctcccactt	tttctcacc	aggctggagt	acagtgaata	66600
tttagtgctt	ttaaatatta	ttcattacca	agccaagtag	cgtactattt	tttccttgta	66660
cagccttttt	ttcccctaaa	taattgtgtt	attttttcat	ttgcttcatt	ttcatcttat	66720
cgattgataa	ttcttctggt	ccatttcagt	ctaatttcca	catggccaaa	ccaatctaaa	66780
aacctgttaa	ctcaggtctt	tttttttttt	ttttttccca	gaataccttt	tcctggaacc	66840
accttcttct	tttcattcga	gaccaaatgc	tttccatgcc	tgtgtcaaag	tggttggtct	66900
ctaactctgtc	ttcaactacca	tcataatttc	cttttccttt	tttcctttct	ttattatgta	66960
tctcctgtct	tttaaattcc	atattttcca	ctttcatgat	ttattttctc	attttgatgg	67020
ctacatcttt	cagtaactga	gaaaagggtg	tgtgagggaa	tattttgaga	tgctgtagtt	67080
ttgaaaatat	ctttattctg	ctgtcttctc	attaataatt	tgactggatg	tcaagattta	67140
agttggaaac	tattttccat	gagtattttg	aaggcattat	ctattgtctt	ctagcttcca	67200
gtgctgcttt	gtattagaga	cagaagatca	ggaatcagga	tagcattgga	cttcttattt	67260
ctattttccca	aagctttttg	gatcttctct	ttattcttat	cttctgatat	tttataatga	67320
tatgcttgat	gcatttactt	tttcttttct	tcttccctct	tgtacatttc	tttctttttt	67380
taggtggggg	caggggggaa	ggagtctcac	tcttctcacc	caggctagag	tgcaagtgtg	67440
caacctcagc	tcactgcaac	ctccacctct	gggggttcaag	cgattctcct	gcctcagcct	67500
cccgagtagc	tgggattaca	ggcatctgcc	accacgcca	gctaattttt	gtatttttag	67560
tagagacggg	gtttcaggct	ggctctgaac	tcctgacctc	aggtgatcog	ccctcctcag	67620
cctcccaaag	tgctgggatt	agaggcatga	gccactgcgc	ccagccctc	ttggaccttt	67680
ctagatattc	atccttcaat	tctgagaaag	gttctataat	tctttaattt	aaaaaaagtt	67740
tttttgtttc	atctttctca	cttttttttt	tttttttttt	ttttaagaca	cagtcttgct	67800
ctgtcacc	ggctggagtg	cagtgggtgcg	atctcagctc	actgcgccct	tcgcctcctg	67860

ggttcgagtg	attctcctgc	ctcagcctcc	caagcagctg	ggattacagg	tgcccaccac	67920
caccatgccc	ggctaattgt	tgtatTTTTA	gtagagacgg	agtttcacca	tgttggccag	67980
gctggtttca	aactcctgac	ctcaggtgat	ccacctgctt	tggcctccca	gagtgtctggg	68040
attacaggca	tgagccaccg	tgcccagccc	tattctgatt	tcatagatgc	agtgtctttt	68100
atgtctctga	gaggTTTTgt	caagTTTTct	tttgtttact	ttattatctg	tttcttcaag	68160
attcctcctc	ttttctTTTT	gccttaagct	ttttcatatt	ggaagcttct	ctcaaagtgc	68220
tggttattct	ggttgccac	ttatattacg	tggaacaata	aaaaagttga	ttaggactct	68280
gtgctgtaag	taagtaaatt	gttgatagtg	agagcccatt	aatcacagga	tgatcaagca	68340
gcaagcctac	tgtttatttg	ggtttctcca	aatggctata	gctgtcagtc	ttttttttct	68400
gggggttattt	ggcttctcta	gaaaagaatt	ttctagtctc	ttgtttggag	aatacaagct	68460
tgactattgg	tgttccagaa	aatgggtggg	aggagacgac	tttgttttct	gttaggttga	68520
gccatgtgaa	attactaggt	tttgTTTTgt	tttctgtttt	gttttgagac	agagtctcac	68580
tctgttgccc	aggctggagt	gccttagcat	gatcttggtc	cactgcaacc	tctgcctccc	68640
aggttcaagc	aattctcatg	cctcagcctc	ccaagtagct	gggattatag	gcatgcacca	68700
ccacgcctgg	ctaattTTTT	tatttttagt	agagacgggg	ttttgccatg	ttgaccagtc	68760
tagtctcgaa	ctcctgacct	caggtgatcc	acctgccttg	gcctcccaaa	gtgctgggat	68820
tacaggatatg	agccaccaca	cctgggctga	aattactggt	tttataggtc	aaaaacagtt	68880
gagggatagg	catggtggct	cccacctata	atcccagcac	tttcagaggc	caaggcagga	68940
agattgcttg	agtccaggag	ttcgacacca	gcctgggcaa	tgtagtgaaa	ccccatctat	69000
atttaaaaaa	aaaaaataga	tgaatatctt	tgtttgcaga	taacatgatt	gcatatgtag	69060
aaaatcccaa	agaaccaaca	aaaagagctc	ctagaactaa	taagtgatta	tgacaagggtg	69120
tagaatacaa	agttaatata	caaagtcaca	ttgctttttt	atctaccagc	agtgaacaac	69180
tggaatttga	aattaaaaca	caatactgct	gggtgcagtg	gctcacacct	gtaatcccag	69240
cactttggga	ggccgaggtg	ggcagatcac	ctgatgtcag	gggttcaaga	ccagcctgac	69300
caatatggtg	aaaccccatc	tctactaaaa	atacaaaaat	tagctgggca	tgggtggcggg	69360
tgcttgtaat	ctcagctact	caggaggctg	aggcaggaga	atcgcttgaa	cccaggaggc	69420
agaggttgca	gtgagccaag	atcgcgccat	tgactccag	cctggggggac	agagcaagac	69480
tccgtctcaa	aaaaaaaaaac	acaatacctt	tcatattaac	actaataaaa	tgaaatatgt	69540

The figure consists of eight vertically stacked diagrams, labeled 1 through 8. Each diagram shows a 2D lattice of nodes, represented by dots. Some nodes are highlighted in black, while others are white. The sequence of diagrams illustrates a process where nodes appear and disappear over time. In diagram 1, a few nodes are black. In diagram 2, more nodes appear. In diagram 3, some nodes disappear. In diagram 4, a new pattern emerges. In diagram 5, the pattern continues to evolve. In diagram 6, the pattern becomes more complex. In diagram 7, the pattern is further refined. In diagram 8, the pattern reaches a final state.

tctttgcccc	ctaaagtctg	ggattacagg	tgtgagccac	cacacccagc	tgtttttaat	73020
tattacatta	atztatgatt	atgtgtttcc	tttaaagcta	tgggcagctc	tgtttgggaa	73080
gatgctactg	aaaaaaggat	gctttttaga	tggcaaagag	tacttaaaat	gtctctagaa	73140
tcatagttgt	aaactaaagt	atatatTTTT	tagttgttca	cagggcttag	agctcctgcc	73200
agagggctgt	tactctttgg	tcacactggg	aatgggaaga	caatgctggg	aagggttctc	73260
ttcaaatttg	agttttctgt	tgagatatTT	gggataatat	gaaaaaaaga	aacttttatct	73320
tgtccttgag	tctattatTT	acgacttgct	ttttgctatt	gtacactttt	gttttttttg	73380
tttgtttggt	ttgttttggt	ttgagtgatc	tgggctcacc	gcaacctccg	cctcctgggt	73440
tcaagcgatt	ttcatgcctc	agcctcccaa	gtagctggga	ttacaggcac	acgccaccac	73500
acctgtctag	gttttttatt	tttggcagag	acaggatttc	gccatgttgg	ccaggctggg	73560
gtcgaactcc	tgacctcaag	cagtcctcct	acctgggcct	cctaaagtga	tgggattaca	73620
ggcatgagcc	actgcacctg	gctgatacac	ttttaagttt	ttcagctact	tttcaatgta	73680
gaagtagatg	gaaaaccatg	tacgttatct	tcagtagtgt	gtttttgggt	ggttaaattt	73740
gacagtatga	ttgtcattat	tttttgtaaa	ttaaattttt	acctggaaga	gcttacctta	73800
ctatattgag	tatctttcta	acccctgatt	tttgcttcta	ctatcataat	aactttatTT	73860
aagtaatcag	tatgttatag	cttttttttt	tttttaagta	ttcttttgcc	agaagttttt	73920
atcaggctct	ggatacctct	ttcctctgca	tagtcctcct	ggatggaaga	aacaaagagg	73980
gaaagagtaa	cttttcctta	gatgtttgtc	tttctcaaag	cagttatctt	tgtatatcta	74040
agaagagagg	agaataacac	tgtctctctt	tttttttaaa	tctctctcta	ctcattctct	74100
ctcaggagga	gagtagaaag	aagcacagct	cttcctataa	cctgtcctta	ttactgagaa	74160
aggaacacat	tgattgccat	gtattgggga	ttgtattata	ccttacattt	ttatTTTTat	74220
tttttattaa	tttttttttt	tagacagtct	tactctatca	ccaagctgg	agtgcagtga	74280
tgtgatctcg	gctcactacc	atctcttctt	tctgggttca	agcaattctg	ccacgtcagc	74340
ttcctgagta	gctgggatta	caggcatgcg	ccaccacgcc	tgactaattt	ttgtatTTTT	74400
attagagatg	ggatttcacc	atgttggcc	ggctgggtct	aaactcttgg	cctcaagtga	74460
tgcgcctgcc	tgcgcctccg	aaagtgctgg	gattacaggc	atgagccacc	acacctggcc	74520
tcatagctta	cattttttaga	gaatcttttc	tagtacttaa	atcggtaaat	atggttatct	74580
tttaaattga	atatattgaa	ctaatttaat	atttgctctt	gtgattttta	aaggctaaag	74640
cagtagctgc	agaatcgaat	gcaaccttct	ttaatataag	tgctgcaagt	ttaacttcaa	74700

gacatgccag	cgaatacgtg	gtcccagcaa	atcaggaggc	tgagggtggga	ggatcacttg	76440
agtcaaggag	gttgagggtg	cagtgagcca	cgatcatgcc	aatgcattcc	aacctgggcg	76500
tcagagcaag	accacgtctc	aaaaacaaaa	caaaacaaac	ttttatgggt	gaaagtgttt	76560
tggaacaacat	acttaaacgt	aaatgtgaat	ctctgatgaa	agaacatggt	acctgtaaaa	76620
gtttgaagtg	tcagcatttg	ttgcaccgaa	atccagaggt	gaggccaggt	gtgggtggctc	76680
acgcctgtaa	tcccagcact	ttgggaagcc	aaggctggca	gatcacctga	ggtcaggagt	76740
tcaaaaccag	cctggccaac	attgcaaaac	cccgtctcga	ctaaaaatac	aaaaattagc	76800
cagacttggt	ggtgtgcgcc	tgtaatccca	gctattcggg	aggctgagac	acgagaattg	76860
ctcgaaccca	ggaggcagag	gttgcaagtga	gccgagatgg	caccactgca	ctccagcctg	76920
ggcaactgag	tgagactctg	tctcaaaaaa	aaaaaaaaaa	aaaaatccag	aggtgaatcc	76980
agaggtgate	accacatgat	atccagatag	cctctttcca	tgagaggctc	aaaggataat	77040
tttactgtct	acagttttgc	agcgagagaa	acttgatttt	atcagtacac	caagagcaga	77100
tctatgtctt	cggaacagac	atgagatcag	aattgtctag	ctgctatgaa	cagcatgttc	77160
tctccctgta	cctatagaca	tgtatgggaa	acttatttgt	aaggttgtat	aatgagcagt	77220
gagttaaagc	aaacttgaca	tgttgaccat	agttgttatg	gcattggact	aaagtagcct	77280
ccatcactat	gatagagata	gccttggtgt	tggaccata	gtttttgaag	tctgtttgct	77340
aagactccct	cttccttagg	cagttcttca	aatattatta	tgctttttcc	tgtcatggga	77400
ttcctcttca	tggagtttcc	tttctcagga	acactatctt	ctcctggtta	atttgtactt	77460
attcttcaga	taccagttga	aatgttacaa	ccttccctga	ccctccaaac	tattccccgt	77520
tattattctc	ctagcaccac	ttgcacctca	ttttcatact	cactggagtt	gcaattcata	77580
ttcattgata	ggattatfff	aattgtatct	gatatcactg	tcacctccac	tagaagatgg	77640
tctccatgog	ggcagagact	atcaccatgt	gttcttcact	tcaattttca	gtagttggct	77700
gtgagtaggt	attgaataaa	tatttgtgga	gtaatcataa	tgaggatatag	atattattct	77760
catagctcta	ttttattaat	tagggaatta	cagaattcag	tgatctgctc	aggatctcat	77820
aaccaggaag	tgggaaacta	ggatttgagc	tccagtgagt	gtggcctttc	attaaaaata	77880
ttacagcaac	cattctcttt	tttttttttt	ttttttttga	gatggagtct	tgtctctggag	77940
tgcaagtggcg	tgatctcggc	tactgcagc	ctctgcctcc	cgggttcaag	tgatccccct	78000
gccttagcct	gctgagtagc	tgggactaca	ggcaccgcc	acgacacttg	gctaattttt	78060
gtattttagt	agagacgggg	tttcaccatg	ttggccagga	tggtctcgat	ctcctgacct	78120

[illegible]

ttgtagttgt	gcttttttac	acagatttaa	tttttatatt	tttcttatat	gttcgaacag	79860
ttaacctgct	tatctattat	taaaaaaaaa	aaaacgaaca	ttcacatagt	tcttaccagt	79920
ttacagtgtt	ttttccacac	cgtcttcaaa	atgtaaagtt	tggtcttcaa	tacatcagta	79980
tgctgctaga	tttaaatact	agggaaaaaa	aaatcagaga	agttaataat	attaactgtc	80040
acctccacta	gaagatggtc	tccatgtgga	cagtaatat	tctcttgat	tatctgtgct	80100
aagtaaaatc	ttctgtaagt	ttctttaaat	attttaataa	atcatagtag	ttaaaatggt	80160
ctcaatatcc	taaagtagtt	aaaagtaact	ataaaatagt	acctgttttt	ctgatcacat	80220
tttacttcct	atgtgaaatt	ttacaagtcg	ttactctatt	tattttattga	tttttttttt	80280
aagacagggg	ctgttctgtc	gccagggctg	gagtacagtg	gcgtgatcat	ggctcactgc	80340
agcctcagcc	tcctgggctc	aagtgatcct	cccaccttag	catccaaggt	agctgggact	80400
ataggcacat	gccaccatgc	ccagctaatt	ttaaaaaatt	ctggggggcc	gaatgcgggtg	80460
gctcacacct	gtaatcccag	cactttggga	ggccgaggca	ggcgaatcac	aaggtcagga	80520
gttcgagacc	agcctggcta	acatggtgaa	accctgtctc	tactaaaaat	acaaaaaatt	80580
agcgggggtg	gggtggcaggc	gcctgtaatc	ccacttactc	aggaggctga	ggcaggagag	80640
ttgcttgaac	ctgggaggca	gagattgcag	tgagccgaga	ctccatctca	aaaaaaaaaa	80700
attttttttt	gtagtgacaa	gggtgtcactg	tggtgccagg	gctggtctca	aacttctggg	80760
ctcaagtgat	cctcccattt	cggcctccca	aagtgtcagg	atcacaggca	tgagtactgc	80820
tgcttggtct	tcaagttggt	attaaagcat	gtttaccac	attatgcaca	tggtataatg	80880
gaaagtattg	ttgtggaagt	taggagatag	ggattctagc	ctagcttttt	atttttttgg	80940
gacaaggtct	cactttttcg	ccccaggccg	aagtgcagtt	gtgcgatctc	ggctcactgc	81000
aacctccaac	tctcaggttc	aagcaattct	cccacgtcag	cctcccaggt	agctgggatt	81060
acaggcatgc	gccaccacgc	ccggctaatt	ttttagtttt	tagtagacac	aggggtttcac	81120
catgttggcc	aggctggtct	tgaactcccg	acctcaggtg	atccaccac	cttggcctcc	81180
caaagtgctg	ggattacagg	catgagccac	cgcaccgggc	ctctagcgta	acttttacat	81240
cctgaactga	ccttaagaaa	gtataacttt	aggcctgttt	catctgtaaa	atgttaatgt	81300
cataggagat	gatcttttga	gatttctttc	agctctgata	attttgtgtg	tgtgtgtgtg	81360
tgtgtgtgtg	tgtgagatgg	agtcttgctc	tgctgcccgg	gctggagtgc	agtggtagca	81420
tctcggtcca	ctgcaagctc	cacctcctgg	gttcacgcca	ttctcctgcc	tcagcctccc	81480
gagtagctgg	gactacaggc	gcctgccacc	acgcctggct	aattttttgt	atttttagta	81540

gagacgaggt	ttcacctgt	tagccaggat	ggctctgatc	tcctgacctc	gtgatccgcc	81600
cgctcagcc	tcccaaagt	ctgggattac	aggtctgagc	cgccgcgccc	ggcctagaat	81660
atttgtttcg	attattctaa	atctggtgac	atttcttttg	tttttaagtt	aaatcttcag	81720
taaaaagaat	aaatgccacc	tagaggacag	aaaaatTTTT	acagtagatt	atcacagacc	81780
tcatgactca	ttactttggt	gtataaaatg	gcctttgtat	ggtgtcagca	cctgggaatg	81840
tctcaagggg	tgttcattga	ccttctggac	tatctggaaa	tattttgata	tttattggct	81900
gggcgcggtg	gctcatgcct	gtaattccag	cacctcgga	ggccaagtta	ggtggatcac	81960
ttgaggtcag	gagtttgaga	tcagcctggc	caactaagtt	agttgactat	ttgtgaaatt	82020
tcctctcca	aataggaaag	ataggattgg	aaacattatt	cagaaggaag	aagttttaaa	82080
gaagggcagg	cttaaagact	atctaataa	tttaccaata	tggtaaaacc	ccatctttac	82140
taaaaataca	aaaattagcc	aggcgtggtg	gtgggcgcct	gtaatcccag	ctacctggga	82200
tgttgagggg	tgagaatcac	ttgaaccac	gaggcaaaag	tcacagttag	ccgagattgc	82260
accactgcac	tccagactgg	gctatagagc	aaaactctgt	ctcaaaaaaa	aaaaggaaaa	82320
aaaaaaaaga	aattttgata	tttatgtgag	aatgactttt	cacggtgttc	ttaatagcgc	82380
aagttttggt	taggagagca	cattccaact	tacttgcttc	tataaatata	ccgtgtaatc	82440
tagggcttaa	agaatatgta	caatgtcttt	ttctctcccc	tagtcttccc	cttttctcac	82500
tagttatata	cttgactgaa	gagctatttc	attctcaagt	cttaggaatg	caggggtgaag	82560
caaacacagat	gaatttttaa	agcacttaac	caggctgtat	gaaatcacag	tctgttgtct	82620
aaaattgtaa	gggacggtta	gtagtactct	cccctttctc	aaaccaaata	tttggttggt	82680
ttaaggaagg	gaaattaaat	tcctgtgtgc	tagattttca	acataaaatt	taaaaaactg	82740
gaataatggt	gcattttatg	tgtataacag	tataatgctt	tgtttttaggt	gggagaagga	82800
gagaaattgg	tgagggctct	ttttgctgtg	gctcgagaac	ttcaaccttc	tataattttt	82860
ataggtaaga	acatattttc	caactaagtt	attgactatt	tgtgaaattt	ccctctccaa	82920
ataggaaaga	tacgattgga	aacattattc	agaaggaaga	agtttttaaag	aagggcaagc	82980
ttaaagacta	tctaataaat	ttagtaggac	ccactatatt	aataagtagt	aaactagatt	83040
aatctcagat	gactcacata	gcttggtctt	taattaaagt	cttataacttg	tatttctctc	83100
agatgaagtt	gatagccttt	tgtgtgaaag	aagagaaggg	gagcacgatg	ctagtagacg	83160
cctaaaaact	gaattttctaa	tagaatttga	tggtgtaagt	gttgattatg	atatttttaa	83220

tgtggcagca	ttttagtata	ttttcctatt	aaatggccaa	ggttaaaaaat	acaaatatct	83280
ttatatattgt	tattactttt	ctaaatgaat	tgaaaaaaga	ttttttgctt	gtaggtacag	83340
tctgctggag	atgacagagt	acttgtaatg	ggtgcaacta	ataggccaca	agagcttgat	83400
gaggctgttc	tcaggtaggg	agatttatat	ggaaatacat	gcatttatta	cagacaatat	83460
ttactcatgt	gtccatctta	catattattt	ccttactctc	agttttaaga	ctaaattcac	83520
tattttcttc	cagtactatc	tctagcctct	tgttaccaac	tacataagga	ttttgaggtc	83580
aacagcttga	tatcaggaga	attagtctag	taaaggttta	gttacggttt	taggcaaatt	83640
ggtcagttct	tctgggttgc	attaaattat	ctttaaactt	gaaaattgat	tcttaccgtc	83700
ttttattggg	cttcatgaga	aaataatcaa	atcagggttc	atggaaaatg	tgttttctgt	83760
tcttggtgtg	ttgagcagtt	tattcccttc	atactggaag	gacggtttgg	gtcagcataa	83820
aattcttggg	catcctttcc	caaggaattt	acaatattgc	tccattgttt	ctagattggc	83880
tgtagtagtg	taaaagtctg	aatccaaatt	gattttttca	aatttgtaaa	ataacttgat	83940
ttttttctct	ctttactgga	tgggttacgt	gggtatatgt	caccaggtta	gtgagcatag	84000
ggtgcagtag	gtagtttttc	aaccacacac	tcactgcttt	cttcccccat	gtagtagtcc	84060
gcagtgtgta	ttgttcccat	gtttatgtcc	atgtgtactc	aatttttagc	tccccactta	84120
tgagaacatg	agctatttgg	tttactgttc	ccacattaat	tggcgtagta	ttaaggcctc	84180
cacctccatc	catggttgctg	caaaggacat	tatttcattc	ttttttatgg	ctgcgtagta	84240
catagtattc	catggtgtgt	acgtaccaca	ttttctttat	ccagtccacc	attgatggac	84300
acttagattg	attctatgtc	tgctgtcatg	aatagtacag	caataaacat	gaaatgcatg	84360
tctttttggg	ataatgatct	attttccttt	gggtatatac	ccagtaatgg	gatttcaggg	84420
tcaaaggtag	atttgtttta	tgttttttga	gaaatcttca	aactgctttc	tacaatgact	84480
gaactttaca	ttcctaccag	cagtacatac	gctccacaac	ctcaccaaca	tctgttattt	84540
ttttactttt	tcccgcctaat	ctgtagacaa	tatgggattt	tttttgccat	tttattaata	84600
gccattctga	ctggtgtgag	atatctcatt	gtgattttga	tttgcatttt	tctgatgggt	84660
agtgatgatg	agcatttttt	catgtttggt	gtccccctgt	atgtcttttg	agaagtgtct	84720
attcatttcc	tttttaccba	ttttttaata	gggttggtgtg	tttttagctt	gttcaattgt	84780
ttgaattcct	tatagattct	ggatatcaga	ccattgttag	atgcagttct	gtagattgtc	84840
tgtttactct	gttgatgggt	ctttttgctg	tgagaagct	cttagtttta	attaggtccc	84900
acttgtcaat	ttttattttt	gttgcaattg	tgtttgggga	cttagccaaa	atttctttct	84960

gaggctgatg	tcaagaagtg	tatttcctag	ttttcttcta	gcatttttat	actttgagat	85020
cttacattta	aatctttaat	ccaccttgag	ttaatTTTTg	tatatggtga	aaagtaaggg	85080
tccagtttca	tttttctaca	tatggctagc	cagtaatccc	catgccattt	attgaatagg	85140
gagtcctttt	cctcattgct	tgtttttgtc	agtcttatca	aagatcagat	ggttataggt	85200
gtgtggcttt	atttctcaat	tttctgtcct	gttccttttg	tttgtgtgtc	tgtttttata	85260
ccagaattat	gctgtttggg	ctgctgtgta	gttcggttta	aagtcaggta	acgtgatact	85320
tacagctttg	ttctttttgc	ttaggattgc	tttgtctatt	tggactcttt	ttttgcttcc	85380
atatgaattt	tagaatattt	ttttctaatt	ttgtgaaaaa	caacattgat	agtttgatag	85440
aaataccatt	gaatctgtaa	attgcttttg	gcagtatagt	cattttacga	ccagataaat	85500
caggcaagag	aaggaaacaa	aaggcatcca	aataggaaaa	gaagtcatca	tactctcact	85560
cttcactggc	aatatgattc	tatactttga	aaaccctaaa	aattccgtca	aaaggctact	85620
agaactgata	aatgatttta	ccccagatt	caggatacaa	aatcagtgtg	caaaaaatta	85680
gtaccatttc	tatacgctaa	taatgtccag	gctaagagtc	aatcaagaa	cacagtccca	85740
tttacaatag	ccacaaagaa	aatgaaataa	tggctgggca	cagtggctca	tgctgtgaat	85800
cccagcactt	tgggaggctg	aggcagggtg	atcacctgag	gtcaggagtt	cgagaccagc	85860
ctgaccaaca	tggtgaaacc	ccgtctctac	taaaaataca	aaaaattagc	caggcgtggt	85920
ggtgggcgcc	tgtaatccca	tctgtctggg	aggctgaggc	aggagaattg	cttgaaccca	85980
ggaggcatag	gttgcagtga	gccgagatcg	tgccactgca	ctgcagcctg	ggtgtgaaag	86040
agcaagaatc	cgttttccaaa	attaaaaaaaa	aaaaaagaaa	aaaaaaaaaga	caatggagta	86100
cctaggaata	cggctgatga	aggaggtgaa	agaatctgta	tgaggagaag	tgtaaagcac	86160
tgctaaaata	aatcagaggt	gacacaaata	aatggaaaaa	cactccatgc	tcatggattg	86220
gaagaatcaa	tatcgttaaa	atgggaaatt	tatcttaatt	tttttccta	gttcattttt	86280
ctgttttctt	ttgcacataa	attatgtgtt	atgttgogtc	cttttgtttt	ccctatctgt	86340
aattttcttc	ttcatccttt	taaactctta	ggtctttttg	ttttgttctg	ttttcccaag	86400
cctgttgccct	ttattttctta	ctgggttagt	atctaatactc	ctttctgcta	tatcttttgt	86460
aactcttctt	tccttggtgg	ttttttatct	attactttcc	tgacatctgt	tattctactt	86520
ttcatcttct	gtttttatttc	tttgagcatc	tgtatgttat	ttccataagc	tcttacattt	86580
cactgggctc	ttgtttcata	gaaaacaatt	ttttggcaag	gacagaccac	ttgaggccag	86640

gagtttgaga	ccagcctggc	caacatagta	caaccctgtc	tctgctaaaa	atacaaaaaa	86700
ttagctgggt	gtggtcacac	acacctgtaa	tcccagctac	tcaggaggct	gataatcgct	86760
tgaaccagc	aggcagaggt	tgcagtgagc	caagacggcg	gcactgtact	ctagcctggg	86820
ggacagagca	aggctctgtc	tcaaaaaaaa	aaagaaaaaa	ttttttggtc	aggcgtcatg	86880
gtgggtcaca	cctgtaatcc	cagcactttg	ggaggccaga	tcacttgagg	ccgggaattc	86940
cagaccagcc	tcgccaacgt	ggcaaaaccc	catatctata	aaaaaaactt	aaggataaaa	87000
aaaattaaat	ttttaaaaaa	ttaaattatt	ttatttcgtg	gaaagatact	tggtcataat	87060
tttagtttgc	tctatggcaa	aatcttctca	tgcattgttct	tcattttatc	tagtttatgt	87120
ttattgtgtt	ttgtttgctc	ttttttgggg	gcagggtttg	ttgttggtgt	tttttttttt	87180
ttttttttta	atgagccctg	tcctgggttt	tccttttttg	cttattattt	ttctttgaat	87240
aaaaggagtt	ttcttggaat	acctagtttt	ggaaggatat	tgtaggggaa	gggataggta	87300
tcttgagtag	tccatgtata	cactaaaacta	gcattgcagcc	tccttcactt	acagtgaagc	87360
cctacctaat	aatgaagtat	atgtgtatgt	aaataagaga	tagggttgat	ttattcttct	87420
actaaaaaat	attgggtcact	gcaagacatt	gtcttcccgc	cagactgtta	gatgccttct	87480
tataaagata	attcgtctta	gagtttcttt	gttcattctc	accttccttg	ccattcctat	87540
gctatcaatt	agggtttata	tggtctgcag	agccaggctt	catatgcttg	tttaaagtgt	87600
gagggactgg	ccttattacc	ctttagggtg	tgctatctct	ttttgggagg	aaaactgtac	87660
tctgacagtc	atgtccatga	tctttagcgg	cagatccttt	gtgtgtcttc	ctttgtcagt	87720
gtaattttca	ctgccctagg	cagttcacct	ttatgtattt	tagttccata	aatgttacct	87780
agtttggtga	tgaagtgtat	gtttctgttt	ctattatcct	tgctgatttt	agtttttttc	87840
agattcaaga	agacagatct	acttatatca	gatattttgt	ttgttttaaa	gcattactgc	87900
tattaaaaaac	atatacatat	ataaatgatt	aattctaaca	ataccgaatt	aaggatttta	87960
aacattaact	agagttttta	aagagtaaca	aaaatatatt	tacattgata	actacaaaaa	88020
tgcatccag	tgccctgaat	attatatatt	aaaagctttt	cctgtcattt	gctgtttcag	88080
ctttaaattc	aaaattatat	ttctaaaagt	gctggatttt	tttttttagg	cgtttcatca	88140
aacgggtata	tgtgtcttta	ccaaatgagg	aggatatgtat	ctgtgtttga	attttttttg	88200
ttttagagca	gaaacaagaa	ctaccatctt	gacaatatta	agtcttccaa	tccatgggtac	88260
agctactttg	gaaaacagtt	tagtggtttc	ttaaaaagat	gtacataagc	ttactgtcag	88320
agccagcaat	ttcactccta	agaatttacc	ccagagaaac	aaaaatgtat	gtccacacaa	88380

agacttggtac	aagaattttt	atagcagcaa	tattaataat	agccagaact	acaaatgatc	88440
taaaatgttt	attttggtga	ataaacaaat	gtggtatatt	catttaatgg	atactgttat	88500
ccattataca	tgctgcatac	tcacttcata	tattaagttt	tatgaatgaa	actccaaaac	88560
attataagtg	aaagaagcaa	gatgatatat	gtcgcatatc	ataggattct	gtttatatgc	88620
agaaaacgca	aattttat	ttattttatt	at	gatggagtct	tgctctgtcg	88680
ctcaggctgg	agtgcagtgg	cgtgatctcg	gtcactgca	acctcacct	cctgggttca	88740
agtgattctc	ctgcctcagc	ctcctgagta	gctgggactg	caggcatgtg	ccaccaggcc	88800
tggtctaattt	tttgtgtttt	tagtagagac	gggtttcact	gtgttagcca	ggatgggtgtg	88860
atccgccctc	ctcagcctcc	caaagtgctg	ggattacagg	catgagtcac	cgtgcccggc	88920
cagaaatcga	aaatttctag	aaacgtaaag	cagatcagtg	ggttgtcagg	gtgagagttg	88980
ggaatgagga	ttaactgcaa	acaggtatga	gtgaactaaa	aagtgttcta	aaactggatt	89040
gtggtgatga	ttgcacatct	ataaatgtat	taaaactcat	tgaattatat	acttacaatg	89100
ggtgaaattc	atagtagatg	aattcatacc	tctataaaac	tggttttttg	cagcaaaata	89160
tttattgcta	tgtttttatt	caagtgtggt	gaattattgc	tatgttttta	tccagggtgtg	89220
gtgtctcacg	cctataatcc	caacactttg	ggaggctgag	gtgggaggat	cgttgagcc	89280
aggaattcaa	gaccagcctg	ggcaacatag	ggagaccctg	tctctacaaa	aactttttta	89340
gaaattaacc	aggcatggtg	gcacgtgcct	ttgggtccag	ttactcgga	ggctgagatg	89400
ggaggattgc	ttgaaccag	gatatcgagg	ctatagtgag	ctatgattgt	gcctgcact	89460
ccagcctggg	taacagcaca	agaccctgtc	tcaatataaa	aaaagaaaag	aatcattaat	89520
tctgaaatta	gactgaatga	tcatttttta	atatttttca	gacaagacta	cttttgctta	89580
aaaatctgtt	atgtaaacaa	ggaagtccat	tgacccaaaa	agaactagca	caacttgcta	89640
ggtgagtaat	ttggatttg	tttatcttac	agcttttatt	tattttttgt	aaataattct	89700
ttttttagtt	atttaaagta	atcttaagta	tgaaatgagt	aattcattga	tcagaagact	89760
ttctctcatc	ctctacctcc	tttgctttt	tcttacctct	tgttcttata	tatatatgtg	89820
gaatctaaat	ttcactggct	atgtcctttg	caagacgtga	tctaattgatg	atagtatatt	89880
ctttttataa	atttattaaa	acctgtaagt	ggtattaaag	taattttaa	atttacacct	89940
taggcaacat	ggtaaaaccc	catctctaca	aaaaataaaa	tgagtcagga	atggtggcac	90000
atacctgtag	tcccagctac	ttggggaggtt	gatgtgggag	gatcgcttca	gcccgggagg	90060

tggaggttga	gcaacagagt	gagaccctgt	ctcaaaaaaa	attttaaaca	tttcatattg	90120
cacactaaca	cagatatattt	tatataaact	tgttcttggc	atatgctgaa	tacttttcag	90180
ctcttttaaat	aattttaatt	atttgataat	tttaataatt	ttaatttcag	ctatttgaat	90240
aatcttaaac	tttatgaaca	gtgtgaagta	gaaactgagt	ataggagatg	cgctcatgaa	90300
ttttattccc	tgaaaaatgt	cataaagtaa	tataaagtat	tcaatgagta	tgtgagtatt	90360
gaataatttt	aacaagaact	ctgagttccc	agatacctat	gacaaaagaa	caaattactt	90420
ggaagtctct	tctactccag	acctttctgt	ctatccccct	atttaaaaaa	atgtacattg	90480
attgtccctt	atcccatcaa	gagagtgttc	tttgccaaga	taaacgctaa	ttagagaagt	90540
aattgtgttt	atgtgtcagt	tggccaccag	tagttttag	atgttacctt	tccaggtgac	90600
atgtttacaa	tttaatgaag	cctctggaga	tagtatgcct	taatctaaag	gaatggttct	90660
aatgattaaa	atttgaggca	ttactacatt	tgttttcagc	aaatgggcct	ttctattccc	90720
ttaaaatcaa	accctcataa	ctatgggttt	gctgacaaag	ggaaactagt	ttttactctt	90780
attatatgga	gatttttaaat	ctgcagagat	gaaaataatt	tgctgtttca	gctgggtgcg	90840
gtggctcatg	cctgtaatct	cactactttg	ggaggctgtg	gcgggtggat	cacctgaggt	90900
caggagtctg	agaccagcct	gaccaacatg	gcgaaacgcc	atctatacta	aatacaaaaa	90960
aaaaattagc	caggtgcggt	ggcgcatgcc	tataatccca	gctacctgag	aggctgaggt	91020
aggagaatcg	ctccaggagg	tggagatcac	agtgagctga	gatcatgcca	ttgcactcca	91080
gcctgggcaa	caagagcaaa	actccatctc	aaaaaaaagc	gggaggggaa	ataatttgct	91140
gtttcttctc	tcccttctct	agaatgactg	atggatactc	aggaagtgac	ctaacagctt	91200
tggcaaaaaga	tgcagcactg	ggtcctatcc	gaggtaggta	tacaagagct	taaaacattt	91260
agaactattt	attataccac	cttagaagtt	taagaagtcc	aaaaaaatct	accaagagat	91320
tttttttttc	ttttggagac	agggctctcac	tctgttgccc	aggctgtagt	gcagtggcac	91380
aatcatagtt	cactgcagcc	tcgaaccccc	caggctcagg	tgattctccc	accagcctt	91440
cccagtagct	gggactacag	gcatgcacca	tcatgccccag	ataattattt	tattttattt	91500
cttgtagaga	cagggtttca	ccatgttgcc	caagctggtc	ttgaactccc	gggctcaagt	91560
gagcagcctg	ccttggcctc	ccaaagtgtc	gggattacag	gcatgagcca	ccatgccccag	91620
ccttaaaacc	attcttagct	cacagatcat	acaaaaacag	tctatgagcc	agtagactgt	91680
gaccctgtt	ctaggcaaga	ggtttcattt	cttgatagta	agtaagcagc	agaaaactca	91740
aaaagaaaac	tttaaacaaa	aaagaataat	gtaaggaaaa	tgttcttatc	tgttttctgg	91800

tgaggtatag	tattttacia	tgatattttc	tttgtcttct	atattgtaag	acatatataa	93540
gacatacata	tatgaatgtg	tgtgtgtgtg	gttttttttt	tttttttttt	ttttttgaga	93600
cagtctggca	tgatatcggc	tactgcaac	ctccacctcc	tggattcaag	ctattctcct	93660
gcctcagcct	cccaagtagc	tacaggcatg	taccaccaag	cccagctaag	ttttgtattt	93720
ttagtagaga	gggcatttca	ccatgttggc	cagtctggcc	tcagactcct	gacctcaagt	93780
gatccacca	actcggcctc	ccaaagtgtc	gggattacag	gcttgagccg	ctgcatctgg	93840
ctgttttggt	tgtttggttg	tttggttggt	tgtttgagac	ggagtctcac	tctgtcgccc	93900
agactggagt	gcagtgggtg	gatctcagct	cactgcaagc	tctgcctccc	gggttcacgc	93960
cattctcctg	cctcagcctc	cctagtagct	gggactacag	gcgcccacca	ccacgcctgg	94020
ctaatttttt	gtacttttag	tagagacggg	gtttcacctg	gttagccagg	atgggtctctg	94080
tctcctgac	tcgtgatctg	cctgcctcag	cctcccaaag	tgctgggatt	acaggcggtga	94140
gccactgcgc	ccggcctttt	tttttttttt	tttttgagac	acagggtctt	gctctgttct	94200
caggctggag	tgcaagtggg	caatcgtggc	tgactgtagc	cttgacctct	tggtctcaag	94260
cagtcctctt	gtctcaaact	tttttagtagc	tgggactaca	ggcaagtgcc	acaacgcccc	94320
actaattttt	taattttctg	tagagatggt	gtctctccct	gttgcccagg	gtggtctcag	94380
actcttgccc	tcaagtaatc	ctcctgcctc	agtctctcag	agtgttggg	actaattttt	94440
tttaatggat	ttttctagtt	gattagaagc	tagaaaaaat	taactttgct	tcattttcaaa	94500
atatggaaac	ctatgaaata	gtcaaaaattt	tttttggtta	tagacaaaac	atatatacat	94560
aaagtcatth	tctatcaaht	catattacat	aaataccaaa	cagaagaaaa	ttacaggccg	94620
ggtgcggtgg	ctcacgcctg	taatcccagc	actttgggag	gccgaggtgg	gcagatcact	94680
tgaggtcagg	agttcaagac	cagcctggcc	aacatggtga	aaccctgtct	ctatcaaaaa	94740
tacaaaaaaa	ttaggccggg	ggcgggtggc	cacacctgta	atcccagcac	tttgggaggc	94800
cgaggcgggc	ggatcacgag	gtcaggagat	cgagacctg	ctggctaaca	tggtgaaacc	94860
ctgtctccac	taaaaaatac	aaaaaaatta	ccggggcggtg	gtggcgggca	cctgtagtcc	94920
cagctacttg	ggaggctgag	gcaggagaat	ggcgtgaacc	tgggaagtgg	agcttgacgt	94980
gagccgagat	cgtgccactg	cactccagcc	tgggcgacag	agcaagactc	catctcaaaa	95040
aaaaaaaaaa	atagctgggc	atggtggcac	gtgcctgtgg	tcctggctac	ttgggaagct	95100
gaggcaggag	aatcacttga	acttgggagg	cagaggttgc	agtgagccaa	gattgcccc	95160
ttgcactcta	gcctgggcga	cagagtgaga	cccagtctca	aaaaaaaaaa	aattacatga	95220

aacttgtagg	ggttttatata	ataaaaaataa	cctataacta	tacgttaatc	ctttaataat	95280
gttctatttt	gtttttcaat	aacctttatt	tttttaatta	atttcctttt	tttgagaagg	95340
actctggctc	tgtcaccag	gctgcagtgc	agtggcacca	tcttggtca	ctgcagcctc	95400
cacctccag	gttcaagcga	ttctectgcc	tcagcctcct	gcgtagctga	gattacaggc	95460
atgggccacc	acgcccagct	aatttttttt	ttttttgtat	ttttagtaga	gatgggggtt	95520
cactatgttg	gccaggtggg	tctcgaactc	ctgacctcag	gcgatcgcca	gcctcggcct	95580
ccctaagtgc	aaggattaca	ggcgtgagcc	accatgcctg	gccaatttat	taccgtttct	95640
taatattgaa	gacctcctta	tgagatgctg	aagcatttca	tttgaaaaaa	gttgcatgta	95700
aaatataaaa	ttgggtatgc	agaatggtca	taactctgtg	agcaaaattt	tgaaatcagg	95760
cttgaattac	agtcagcctt	ttctatccgt	ggattctgca	ttcatggatt	caaccaactg	95820
cagatcaaaa	atacttgggg	gaaaaaagca	tctgtgttga	acatgtacag	actttttattc	95880
ttgttattat	tccttaaaca	atatagtata	aaaactattt	acattgcatt	tacattgtat	95940
taggtgttat	aagtaatcta	gagatgattt	aaagtataca	ggaggattgt	gtagggttga	96000
tgcaataact	gcaccatttt	acataaggga	ctttaagcgt	ttgcggattt	tgctatctat	96060
gggggaacca	atttccgaga	gatactgagg	gacagctgta	tatttgtaac	ttatttttta	96120
tttcccta	tgcagcagct	gttgagggga	cagtgaactg	ttaacacaga	taacaagtgt	96180
attcaagtac	attttggagt	ttgtttgttt	ttttttttca	ggctttttatt	cgctgtcgg	96240
atgaggcacc	atacttgaat	tttttttttt	aagaaagctt	ttagttttct	ttctttcttt	96300
tttttttttt	gtgagatgga	gtcttgctct	gtttctaggc	tggagtggag	tgcagtgggtg	96360
tgatctcagc	tactgcaac	ctccgcctac	tgggttcaag	tgattctcct	gcctcagcct	96420
cccagtagc	tgggactaca	ggtgcacgcc	accacgcccc	gctaattttt	tttttttttt	96480
tttttttttg	tatttgtagt	agagacagag	tttcacatg	ttggccagga	tggtctcgat	96540
ctcctgacct	tgggatccgc	ccaccttggc	ctcccaaagt	gctagggatt	acaggcatga	96600
gccaccacgc	ccggtgaaa	gcttttagtt	ttctaactta	tttaatttaa	tttaatttaa	96660
tttattttta	ttttattttt	tgggacagtg	tctcacttgg	ttgccaggc	tggagtgcag	96720
tgggtacaatc	atggctcact	gcaccctcta	gctcctgggc	tcaggcaatc	ctcctgcctc	96780
agccttttga	gttgctggga	ctacgggcat	gtaccaccac	actcagctaa	atttttaatt	96840
ttttgtagag	atgggggtctc	actatgttgc	ctaggctggt	ctcagattcc	tcaagcattc	96900

ctcccacttg	cacctcccaa	aatgctggga	ttacaggtgt	gacaccgtgc	cagacttgaa	96960
atttttaatc	ccacacctaa	aaatataatt	ttatccacca	ttttttaaaa	gtcataacat	97020
tatttattaa	aaatttagat	ggtaaaacta	aaaattaaag	cttaataaaa	ctactgagtg	97080
atttacaagg	aagaatatta	ctggtcacct	ttgtgagcat	cccattaatt	atatatatct	97140
aggttatatt	gttacaatat	ttggttctac	tgtatacttt	tttttttgag	actgagtttc	97200
gctgttggtg	cccaggetgg	agtgcaatgg	tgcaatctcg	gctcactgca	acctccgcct	97260
cctgggttca	agtgattctc	ctgcctcagc	ctcctgagta	gctgggatta	caggcatgcg	97320
ccactatgcc	tggttaattt	tgtgtttttg	tagagacggg	gtttcactat	gttggtcagg	97380
ctgatctgga	actcccaacc	tcaggagatc	cgccacacct	ggcctcccaa	agtgtcggga	97440
ttgcaggtgt	gagccaccat	gcctggctct	actatatact	ttcattcagt	tgtttctttt	97500
taatctagtg	gttttggtat	taataatttg	ataatgacct	ttagctgtta	ttgcttactt	97560
atgagttaat	atttataaag	cacttaaagt	agctggcacg	tagtaaacac	tatgtaaaga	97620
tccattaaat	aactttaaaa	aatataaaac	tgatagtggc	attttattat	agagattaag	97680
gtaatccatt	ctctcatttc	catttatggg	atgagacgta	aacacaagta	gtttgctctc	97740
taaaactgta	tactagaatt	ttatataccc	attatttgat	gcaactttaa	taccaaagtg	97800
tattgtcagt	taccggtgaa	tatatataaa	tttaggtaag	gaaaacccaa	cttggtcgca	97860
cacagtggct	cacgcctgta	atcctaccac	tttgggaggc	cgaggcagtt	ggatcacctg	97920
aggtcttgag	ttcgagacca	gcctggccaa	cgtggcgaaa	ccccgtctcc	actaaaaata	97980
caaaaattag	ccgggcgtgg	tggcacatgc	ctgtaatccc	agctgctcgg	gaggctgagg	98040
aaggagaatt	gcttgaacct	ggggaggtgg	aggttgcagt	gagctgagat	cgtaccactg	98100
cactccagtc	tgtgcgacag	gagcgagact	ccatctccaa	aaaaaaaaaa	aagaaaaacc	98160
caacttatct	tttacagttt	ataatagtag	aagttcaaat	aattggtttg	gaatttctgt	98220
atttttaagg	ttagtactaa	aattgttggt	tataaattgg	ggtacaatat	actttgtttt	98280
ttaagtacct	tgtgtatcta	atttaacttt	aagtccttta	ttattttggt	ttgataagac	98340
aacttttcta	cttatttccc	cccttaactg	aaccagctac	catctgcctt	tttctgtttg	98400
tatacattag	tctcttacgt	taaaatatca	tataagtttc	atatatatac	acatatcaaa	98460
ctatagactt	aaagtacaat	tacatcaaca	tcttttaaaa	ccttaatttc	tggccaggcg	98520
tggttggtca	tgcttataat	cccagcactt	tgggaggctc	aggccattgg	atcccttgag	98580
ctcaggaatt	cgagaccagc	ctggataaca	tggtgaaatc	ctcatctcta	caaaaattac	98640

aaaaagttag	ctgggcgtgg	tggcgcacac	ctatagtcct	agctacttgg	gaggatctct	98700
taagcctggg	aggcagaggt	tgcagtgagc	cgagatcatg	ccactgcact	ccagcctggg	98760
taacagtgag	accctgtctc	aaaacaaaaa	tattaatttc	tataacaaat	aaattttatt	98820
tagtaatatc	atgtttaata	cccatgttac	attcattatt	cttctacact	ttggccttta	98880
ggaagaggca	gtgtgtgtgt	tacagttaaa	taggaaacaa	gtagggttca	tatagtgcct	98940
atggggtttt	ctttgggtgg	gggttgctgg	gtttttgggg	tttttttttt	tgagacaggg	99000
tatcactctg	tctcccaggc	tgaaatacat	tggcatgctc	atggctcact	gcagcctcaa	99060
cctcccagac	tcaagtgatc	ctcccacctc	agcttcacct	gtagctgaga	ctacaggcgt	99120
gcgccaccat	gcctagctaa	attttgtaca	ttttgtagag	atgaggcttt	gccatgttgc	99180
ccaggctggt	ctctaactcc	tgggctcaag	tgatccactt	gcttcagcct	cccaaagttc	99240
tgggattaac	aggtgcgagc	cactgcagcc	agcctatatt	ttactatat	gtttttttct	99300
ttttggctaa	aatttttcag	attagtttac	aagttacaag	tgtaggtgat	atctcatgga	99360
gatatacaa	atgatatgaa	gttagattgg	gtttttaaga	gtagttttta	aaatacggat	99420
aaataccagt	tgttgagatt	ttgtttaaag	ttctttaact	tcttattttt	gggccaaaaat	99480
acaggtatac	gctagaaatg	atttttaaca	caggtcattt	atgccaaaact	gcattttgcc	99540
ttaatttttt	tttttttttt	tttttttgag	agggagtctc	actcattgcc	caagctggag	99600
tgcaatggca	cgatctcggc	tcacttacaa	cctccacctc	ccaggttcaa	gcgattctcc	99660
tgctcagcc	tcctgagtag	ctgggattac	aggcacctgc	ctggctaatt	tttgtatttt	99720
tagtggagac	ggggtttcac	catgttggcc	aggctggtct	cgaactgctg	acctcaggtg	99780
atccacccgc	cttggcctcc	caaagtgcgg	ggattacagg	tgtgagccac	cgtgcttggc	99840
caaaattttt	attaattttt	ctattgcctg	gactctgtga	acctatccat	tttgcttttt	99900
aaaaatactt	aggtgtaaat	atagatattc	attaactcag	cattgtttta	atctatatatt	99960
ccaaaggcaa	tttaaaagat	cagaaaataa	gaccaaatta	atataaaaaat	gcatacttta	100020
ggctgggcaa	agtggctcac	gtctgtaatc	ccagcacttt	gggaggccaa	ggtgggcgga	100080
tcacctgagg	ttaggagttc	aggaccaacc	tggccatcat	ggcgaaaccc	tgtctctact	100140
aaaaatacaa	aaattagctg	ggcatggagg	catgtgcctg	taatcccagc	tactcgggag	100200
gctgagacag	aagaattgct	tgaaccttgg	agggggaggt	tgcatatctg	agtggtgaaa	100260
ttgtgattct	ttttttctct	ttgtctgtat	ttttgaactt	ttctataaat	gatttgtgtt	100320

tgtttttata	ttggaaaaat	attatgcttt	caaagtgtta	tacctatgaa	actaaacaca	100380
agtaataaat	atattagtat	agcattttatt	aagggtttctt	gtgtagcaga	tcaacataga	100440
aaatatat	aaatggctga	cataattttc	taagaataca	tacacgtata	ttttttataa	100500
cattaagaaa	cagcagcatc	attactttta	tccatcattt	cgttaaccac	catataacctg	100560
ttgatcattt	gtattgtcat	gtgcttttta	aaaatctaga	tgagaaatat	tcgattatct	100620
gacttcactg	aatccttgaa	aaaaataaaa	cgcagcgtca	gccctcaaac	tttagaagcg	100680
tacatacggt	ggaacaagga	ctttggagat	accactgttt	aaggaaatac	ctttgtaaac	100740
ctgcagaaca	ttttacttaa	aagaggaaac	acaagatctt	caatgaacgt	catcggtctac	100800
agaaacagcc	taagtttaca	ggacttttta	gagtcctaca	tatttggtgca	ccaaacttga	100860
agatgaacca	gaaaacagac	ttaaacaata	tatacaatgc	aaatgtaatt	ttttgttggt	100920
taaggccttg	ccttgatggt	cacagttatc	ccaatggaca	ctaagttaga	gcacaacaaa	100980
acctgattct	ggtcttcttt	accaatataa	tcataatgta	aataataatt	tgtatattgt	101040
gttgacagatg	aaagtattcc	aggaacagtg	aatggtagaa	gacacaagaa	catttgtttg	101100
tttgtcttct	gatgtttttt	cttaaaatag	taatttctcc	tacttttctt	ttctactggt	101160
gtcttaacta	caggtgattg	gaatgccaaa	cactcttaag	tttattttct	tttttcgttt	101220
tataaattca	gtgtgccaaa	tgaaactttt	ttcctaagta	actgtaatag	gaaaaagt	101280
atgttgagag	tttcttcttc	ataaatctac	agacattaaa	caattgttgt	gttcttttta	101340
ccttttattt	ttctattacc	ttgctaccaa	acagtttaga	tagcaatata	atagcaaaaa	101400
agcaaatatg	gtaaaataga	gaagggttga	agggttgagt	tactctgtca	tataacatgt	101460
agatcagtct	tcattgtgacc	tgcagtattt	ttttttctaa	tgtatttgtc	agaaatctgt	101520
tgtagactgt	taacttcttc	ctgatggaat	ttattttctg	caagaattat	tctgatattt	101580
aagagagcca	atttttaactg	ctgtgaaaat	gtttccagtg	caagagaagg	gaaatactag	101640
gaactaagac	atttctaatt	tattgcttat	tactttctta	attttacagg	ataattataa	101700
gcaagtggaa	ctaccatctt	ttattcttaa	taattattaa	tcccttcaat	gaaactttta	101760
aaaaactgaa	tttttataca	tggcatacat	ttttctagtt	ccttctgctt	gctttattaa	101820
ctcaaaagtt	ctagttctag	tctgttgatc	tgcttttgt	tctcccaaaa	tgtacagtaa	101880
ttccatttgt	ttgtataaat	atgcctggat	tttcattata	aaaatgtcat	tgtaggaggat	101940
agagactcat	atcatggcct	tttaaatatt	gtaataaagg	caaataagata	tttgccctta	102000
gtttactggg	taaaagtttg	tttacagaac	ttttctctgg	tgcttaaatg	atgctatgta	102060

aaatgtcatg	agtggaaaga	atattttgtag	tagtaacaag	aatttttcat	ttaggaaaga	102120
tttcttaggt	tttgaaagaa	tacattaaaa	taaaaaactt	gccctacta	ggtaagaact	102180
ttataatgaa	gacatacatt	cttcttaatt	ttactcttgc	tcttggttaa	gatttggttg	102240
aatatagaag	atgcatgatt	tctgggtttt	tttttttttt	tgagacagag	tttcgctctt	102300
gttgcccagg	ctggagtgca	atggcgcaat	ctcgactcac	cacaacctcc	gcctcccagg	102360
ttcaagcaat	tctcctgcct	cagcctcccg	agtagctggg	attacaggca	tgcgccacta	102420
ccccagctaa	ttttgtattt	ttagtagaga	tggggtttct	ccatgttggt	caggctggtc	102480
ttgaactcct	gacctcaggt	gatccgcctg	cctcggcctc	ccaaagtgct	gggattacag	102540
gcataagcca	ctgcgcccag	ccagaagatg	catgatttct	taggatcata	tgctgtttgt	102600
agccataagg	taaatcatgt	ctcttccaat	catgactttg	gaactccctg	aataataaaa	102660
atgagagttg	agataaatag	gggaaaaaaaa	atttttttca	agccagagct	atgcatatgt	102720
taggtgatgg	gtagtatccc	tttaagggtct	caaacattac	aacatcaatt	atgaaatact	102780
gataacgaaa	ggtagtaatg	aaatatatat	gatgaaaaga	attgagaagt	tctaaattaa	102840
gacatttcag	ttaagctcat	aaaatttcatt	tgtttttcatt	taaaagatta	acgttattga	102900
tacttgata	actggctaata	catattaaag	gactatgtgg	ttccagctca	acttttaata	102960
tattgtctcc	tttaaaacta	tcatgggtat	aattctattg	ggaaagactt	ttagataaca	103020
aagatttcaa	atgttaaaaag	agataaaaagt	cagggttaata	ctatcttaaa	cactgagtca	103080
gaaaatcatt	actgtataga	agttgctttc	ctgatcaagt	ctgaacttca	gctagtgcta	103140
gagaactatt	ttctatgact	taactctaac	caagttttat	tttaagctgt	ttctttgata	103200
gaagggccat	gaaaatagag	taatgatata	gtaggagata	agggattggt	ttgggtctttt	103260
tcaataaaga	tagaagttgc	tgaagttttc	tgaattaata	atgacttaga	ttgtgacctt	103320
ttagattcgg	tgttgagctc	tgtgttgat	tacttcctaa	aagataatgc	ttaaaccatta	103380
agcatttagtg	tgctcttcat	gttaatatgg	cagagttttg	taaactaaat	taaaacttac	103440
tgatatattg	gactttgagc	caagggaaaag	aatgagtact	atctttccag	atatcttaag	103500
ggtaaaaagct	tatttctaaga	cagtctgtcc	attgagaata	ttagatttct	gacttgcaaa	103560
tatgtttgta	ctccagaaga	attagaggaa	aagcagatac	tagaattcta	atttaattac	103620
atatacagcc	gtctttgttt	atagtgtaga	attctttata	ttttgtacaa	aaactaattc	103680
ttttggtaaa	atgaaccatt	tacagttcgg	ttttggactc	tgagtcaaag	gattttcctt	103740

taaatgcttg	tctcaatfff	agtctggtct	tttgtacttt	tcttcagaag	aatgaatta	103800
aagggtagag	ttgcataaag	tgggttttta	tcctaagtga	ttggaaataa	atgataaact	103860
ttatfffgtc	tttactffff	tactfftaaaa	ctffffgata	ttttaggggt	tggagtctga	103920
taatgaagga	gttgtgtgta	ttggactctt	agtaacaatt	ataaacgctt	aacaaaatat	103980
agaaagcaat	gattgggtgg	ctctgcagag	caatcaaaac	aaggtagaaa	ctgcaaagtc	104040
ctataatgga	agagattcag	gctgggtgtg	gtggctcatg	cctgtttggg	aggccagtat	104100
gggagaatcg	cttacctca	ggaattcaag	accagcctgg	ggaatatagg	gagaccctgt	104160
ttctgccaaa	aaaaaaaaaa	aaaaaaaaac	taaactaaaa	attagccagg	tttgggtggc	104220
tacacctgta	gtcccggtta	ctcaagaggc	tgaggttaga	ggatcgctta	agcccgggag	104280
gaggaggttg	cagtgaacca	agatcacacc	accacactcc	agcccagggt	acagagaacc	104340
tgtctcaaaa	aataaaaagt	aaacgagatt	caccttaacc	agcttttacc	catagggcaa	104400
tttccagttt	gtgcagttca	cttgatata	gaattcaggt	agaatgtgac	agttttgaca	104460
ggctgaggaa	tcatttggga	cttctagaac	acctgaaaat	tagaagagaa	atfffgggaa	104520
ggagagtgcc	tcagaaagta	agctctaaaa	tctgcctaca	tattcctttc	aaatccttgg	104580
tggattccaa	aattgtgctg	gtgcagtgtg	attatfftaa	ggaaccaggg	agaaagcaag	104640
ttaaagtcta	aaaaactcag	cagcgattgc	agctgcccaa	ggacagagag	tttgaaattc	104700
aagttccact	aagaaggact	gagtaaatac	ttggtgtttc	ccattgaacc	cccaaatgc	104760
cacgccttaa	ggaagaatga	cagcatccta	gaactaaagg	ctgtgcttca	gcactaagga	104820
caaatggaa	ataaacttac	tgtaatgaag	cttaaatca	aatctcacag	catcaagggt	104880
atccattagt	aattftaaac	cagaacaaaa	cacagcattc	ttgagaagaa	aacaattcag	104940
tcatcacagc	atatatccaa	agctcagtag	ataataaaca	gttactaagc	atgcaaagaa	105000
aaattatgtg	atcctffftga	ctcaacttcc	ctgatggaaa	aaaaaaagtg	gtccatacca	105060
tactcaacag	aaaatataca	gaagcagatg	cacaagttag	ctctatgaga	gtgattfftaa	105120
aatagcgfta	aaatatgtta	aaagaattta	caagaagaga	taatgggtga	atagataggg	105180
agtttcagca	gaaaaataag	aattgaaaac	ccaacagagg	ccaggcaccg	tgcttcacgc	105240
ccgtaatcac	agcactfftg	gaggccaagg	cagggtggatc	acctaaaggtc	aggagttcaa	105300
gaccagcctg	accaacatgg	tgaaaccggg	tctctactaa	aaatacaaaa	ctagccggggc	105360
atgggtgatg	catgcccata	atcctagcta	cttggggaggc	tgaagcagga	gaattgcttg	105420
aacctgggag	gtagaggttg	cagtgaacca	agatcatgcc	attgcactcc	agcctgggca	105480

agaagagcaa	aactccatct	caaaaaataa	taataataat	tggctgggcg	cggaggctca	105540
cgctgtaat	cccagcactt	tgggaggccg	agacaggcag	atcacaaggt	caggagattg	105600
agaccatcct	ggctaacatg	gtgaaacccc	gtctctacta	aaaatacaaa	aaattagcct	105660
ggcatggtgg	cgggcacctg	tagtcccagc	tactcgggag	gctgaagcag	gagaatggca	105720
tgaccccagg	aggcggaact	tgcaatgagc	ccagatcgcg	ccactgcact	ccagcctggg	105780
caaccgtctc	aaaaataata	ataataataa	ttagaaaata	ataaaaaata	acccactgga	105840
aagtcttgta	atgaaaacta	cagtatctga	aattaaaaat	ttaaattgat	ggacttaaca	105900
gtaaactagg	cacaacagga	aagactgaaa	aagacaagta	agtatcaaac	caaagcagag	105960
agagaaaagaa	aaaagaaata	gcagaggttt	gagacctttg	gaacattatc	agttcctgta	106020
at ttgaagta	aagaaggaaa	ggaagagaat	ggggcaagag	aaatctttga	agcaatgatg	106080
cgtgaaat tt	tccccaaagt	gttgaaagac	atcaacgtac	agatctaaga	agttcactga	106140
atccccaaagc	agaataaata	caacaaagac	tacatctggg	tacatcacat	tgccaaaaaa	106200
aaaaaaaaaaaa	aaaatcttaa	aagcaatcaa	ggtggggagt	gggaaggagt	agcattacat	106260
tcagataaac	aagagtgatg	gctaacctca	tcagaaatga	tggaattaag	aaaatgatgg	106320
aatgacatct	ttaaggtgca	aagatgaaac	aaaggttaac	ttaaaactct	gtatccagtt	106380
aaaatatcct	tcaagatg tt	aaggcaatgg	ccaggctcag	tggcttatgc	ctgtaatccc	106440
aacactttgg	ggggctgaga	caggagaatc	ccttgaggcc	atgagtgacc	agcctgggca	106500
acctagttag	aaccccatca	aaaaataaaa	ttagctgggc	atgggctgtg	gtcgtgctac	106560
tgcactccag	cctgggggac	agagtcccta	tctcttgaaa	aacttaagac	aaaatctttt	106620
ccagatgaaa	acagaatttg	ttgctggcag	actgacttta	tacggaatgc	tttaagaaat	106680
tcttcaggct	gaaggaaaaat	ataaatctgg	agtgaaaaac	gtagtatctg	aaattaaaaat	106740
tttccttgga	tggacttatc	agtaaatgga	tttcccaatg	ggaacctgta	tctccaggat	106800
ggaattaaga	at tttgaaat	gataaacatg	ttggtaaata	agaaagacat	ttttcttttt	106860
tat ttttcta	aatatcattg	ctgtttaaag	caaacaaaaa	tagtattgca	gtgtttataa	106920
tatatgtata	agtaaaatat	aggacagaaa	tagcatgaag	ggtcagagag	ggcaataaat	106980
gtaattatat	tgagataata	ttactatatt	gtacacaaag	tgagatagca	ttagtgtaag	107040
gtagattgtg	ataagaatgc	at tttgtaat	ctctcacccc	atcactgggg	tgca gtagat	107100
ctaaaaagcc	attagagaag	ataaaatggg	tcactaagaa	ttatgtgatt	agtcaaatga	107160

aatcaaggaa	agaaaaacag	agcaaagaag	tgtaagataa	aacagtggga	tggaagagct	107220
gaaattaata	atgtcactaa	attgcttacc	ttcctccttc	ctgaactgtt	gtggatTTTT	107280
TTTTTTTcag	ggggTTgagg	TTTTTTTTta	gagacagagt	ctctattgcc	ccaggctgga	107340
cagcagtggc	gtaatcagtc	tcaaaatccc	tgcagtctca	aactcctggg	ctcaagtgat	107400
cctcctgcct	cggcctccca	aaatgctggg	attacaggca	ggcagagcca	ccacaccctg	107460
ccagaactat	tgTTTTaaag	tcattaggtg	gttaagccct	cgagactatg	tgcaagTTtg	107520
atgattcggt	aggagcactc	acaggactca	gcatatagtc	ttattcccaa	caatgattta	107580
ttagagcaaa	aggatatgaa	gcaaattcag	caaagggaaa	aaaggaatta	agtgaagtcc	107640
agaggaaacc	aagcaccagc	ttctaagggt	tctctcctaa	tgaagtcaca	caggatgtca	107700
ctccagcaac	aagtgcagc	atgtgaaagt	gttaagcctc	actgttagac	atagtaagct	107760
tgttagggaa	tagagggaat	ctccccctca	attcaagttc	ccagatgcct	ggcaagggcc	107820
aacettgcc	gcaggacttt	ctggggatag	cagtcccagg	cacgtgtttg	cacagtgggt	107880
cagatcaatg	tgcttacatt	gggtagagg	gacctatgga	agtccaaatt	tgggtgtcag	107940
agacctaata	gggtgaagac	agtgtctaca	atgatggaca	gctaggtatg	aggtgtcagg	108000
gacagattga	ggcagttatt	cacatggtgg	ataggggcaa	cctggaatga	ggaataagct	108060
caagcacaga	aaggggtgtc	catatgggg	aaggggtgtc	gcagagatgg	tagattgggt	108120
gcacaccaaa	gacttgatgg	aataaaagtga	atatactaac	cacagagaag	gtaattataa	108180
atacagaaa	gggggaaact	agaatgaaac	ctggagttca	gcttgaattg	ggtttaagaa	108240
agtgaattca	tggTTTTaaa	tctataaaaa	tagatgaaat	atcggtctgg	cacagtggct	108300
cacgcatata	atcccagcac	tttgggaggc	cgaggcgggc	agatcgcttg	agatcaggag	108360
tttgagacca	gcctggccaa	cactgcgaaa	ccccatctct	agtaaaaata	caaaaattag	108420
ccagggtgtg	tggtacacac	ctgtagtccc	agctacttgg	gaggctgagg	cacgagaatc	108480
acttgaacct	gggaggcaga	ggttgcaagt	ggctgagatc	gcaccactgc	actccagcct	108540
gggcaacaga	gcgagatcct	gcctcaaata	aaaatagagg	acatataagc	atataaatat	108600
acacgtgtgt	atgtgtccaa	atatgtatat	tccctagtct	gtccaccaag	gtggccttgg	108660
agcagttatg	ctccaataat	aatgagcaca	taaagtaccc	atatcttgcc	ttccaaattc	108720
ttcactgtct	tagtctgctt	gggctgcatt	acaaaatacc	atagactggg	cagcttaa	108780
aacagaaatt	tatcctcaca	gttctagaag	cttggaagtc	caagattaaa	gtaccagcca	108840
gtttggtttc	tagtgagggc	tttcttcctg	gcttgcagat	ggccaccttc	tcaccgtgtc	108900

cttgatggc	agacagcaca	agctctctgg	tgtccctttt	taaaagggca	ttaatcccat	108960
catgacagtc	ccatcctcat	tatctcatct	aaccctaggt	acttcccaaa	ggctgaatca	109020
ccaaagacca	tcacattgct	ggtgaaggct	tcaacatatg	aatttgaggg	acacgaatat	109080
tcagtccata	acatcaacta	aaggaaccaa	gactctttga	taaaatggct	aaattcaggg	109140
ctggggcaga	gaaaatacat	gagtgtggaa	cttcttgtgc	cagagagaaa	aagtgcccaa	109200
agattgatga	ggatgaatca	ttgaaatgac	acacagatta	aaagggttcc	cactggacaa	109260
atttgagcat	caaaataagt	aatagtagta	attaattata	acccatcaga	agaaataaac	109320
catgagctca	tgtgaatata	tgaatacaaa	cataaacaaa	ttacaagcat	aatgaggaat	109380
gtgatattta	tatggtttta	aggtacctct	ccaggccggg	tgcagtaact	ctcacctgta	109440
atcccagcac	tttgggaggg	caaggcaggt	agatcacctg	aggtcaggca	tttgagacca	109500
gcctgcacaa	catggtgaaa	ccctgactct	actaaaaata	cataacgcga	gccgggcgtg	109560
gtggcacgtg	tctataatct	gccactgatt	aggtgtgtga	ttttcccaag	caggggataa	109620
tagtagtacc	tatgtcaaag	gctgttatga	ggattaaatg	agctaacaca	taatcgtgct	109680
tttttttttt	tttttttttt	ttgagacaga	gtcttgcact	gtcgcctggg	ctggagtgca	109740
atggcacgat	ctcgcccccac	tgcaacctct	gcctcccagg	ttcaagtgat	tctcctgcct	109800
cagcctcctg	agtagctggg	attacaggt	cctgccacca	cacctggcta	ttttcaatag	109860
agacgggggt	tcactatggt	ggccaggcta	gtctcaaaaa	cctgacctcg	tgatccaccc	109920
gctttggcct	cccaaagtgc	tgggattaca	ggcatgagcc	actgcacccg	gctttttttt	109980
tttttttttg	agatggaatc					110000

<210> 2
 <211> 3263
 <212> DNA
 <213> Homo sapiens

<400> 2						
gctcctgaga	ccggcggggca	cacgggggtc	tgtggccccc	gccgtagcag	tggctgcgcg	60
cgtcgcttgg	ttcccgtcgg	tctgcgggag	gcgggttatg	gcggcggcgg	cagtgagagc	120
tgtgaatgaa	ttctccgggt	ggacgagggg	agaagaaaag	ctccggcggc	gccagcaacc	180
cggtgcctcc	caggcctccg	cccccttgcc	tggcccccgc	ccctcccgcc	gccgggccgg	240
ccccctccgc	cgagtcgcgc	cataagcgga	acctgtacta	tttctcctac	ccgctgtttg	300
taggcttcgc	gctgctgcgt	ttggtcgcct	tccacctggg	gctcctcttc	gtgtggctct	360

gccagcgctt	ctcccgcgcc	ctcatggcag	ccaagaggag	ctccggggcc	gcgccagcac	420
ctgcctcggc	ctcgcccccg	gcgccggtgc	cgggcggcga	ggccgagcgc	gtccgagtct	480
tccacaaaca	ggccttcgag	tacatctcca	ttgccctgcg	catcgatgag	gatgagaaaag	540
caggacagaa	ggagcaagct	gtggaatggt	ataagaaagg	tattgaagaa	ctggaaaaaag	600
gaatagctgt	tatagttaca	ggacaagggtg	aacagtgtga	aagagctaga	cgccttcaag	660
ctaaaatgat	gactaattttg	gttatggcca	aggaccgctt	acaacttcta	gagaagatgc	720
aaccagtttt	gccatttttc	aagtcacaaa	cggacgtcta	taatgacagt	actaacttgg	780
catgccgcaa	tggacatctc	cagtcagaaa	gtggagctgt	tccaaaaaga	aaagaccctt	840
taacacacac	tagtaattca	ctgcctcgtt	caaaaacagt	tatgaaaact	ggatctgcag	900
gcctttcagg	ccaccataga	gcacctagtt	acagtggttt	atccatggtt	tctggagtga	960
aacagggatc	tggctctgct	cctaccactc	ataagggtac	tccgaaaaca	aataggacaa	1020
ataaaccttc	tacccttaca	actgctactc	gtaagaaaaa	agacttgaag	aatttttagga	1080
atgtggacag	caacottgct	aaccttataa	tgaatgaaat	tgtggacaat	ggaacagctg	1140
ttaaatttga	tgatatagct	ggtcaagact	tggcaaaaca	agcattgcaa	gaaattgtta	1200
ttcttccttc	tctgaggcct	gagttgttca	cagggcttag	agctcctgcc	agagggctgt	1260
tactctttgg	tccacctggg	aatgggaaga	caatgctggc	taaagcagta	gctgcagaat	1320
cgaatgcaac	cttctttaat	ataagtgctg	caagtttaac	ttcaaaatac	gtggggagaag	1380
gagagaaatt	ggtgagggct	ctttttgctg	tggctcgaga	acttcaacct	tctataattt	1440
ttatagatga	agttgatagc	cttttgctg	aaagaagaga	aggggagcac	gatgctagta	1500
gacgcctaaa	aactgaattt	ctaatagaat	ttgatggtgt	acagtctgct	ggagatgaca	1560
gagtacttgt	aatgggtgca	actaataggc	cacaagagct	tgatgaggct	gttctcaggc	1620
gtttcatcaa	acgggtatat	gtgtctttac	caaatgagga	gacaagacta	cttttgctta	1680
aaaatctggt	atgtaaacaa	ggaagtccat	tgacccaaaa	agaactagca	caacttgcta	1740
gaatgactga	tggatactca	ggaagtgacc	taacagcttt	ggcaaaagat	gcagcactgg	1800
gtcctatccg	agaactaaaa	ccagaacagg	tgaagaatat	gtctgccagt	gagatgagaa	1860
atattcgatt	atctgaattc	actgaatcct	tgaaaaaaat	aaaacgcagc	gtcagccctc	1920
aaactttaga	agcgtacata	cgttggaaca	aggactttgg	agataccact	gtttaaggaa	1980
atacctttgt	aaacctgcag	aacattttac	ttaaaagagg	aaacacaaga	tcttcaatga	2040

```

acgtcatcgg ctacagaaac agcctaagtt tacaggactt tttagagtct tacatatattg 2100
tgcaccaaac ttgaagatga accagaaaac agacttaaac aaaatataca atgcaaattgt 2160
aattttttgt tgtttaaggc cttgccttga tggtcacagt tatcccaatg gacactaagt 2220
tagagcacaa caaaacctga ttctgggtctt ctttaccaat ataatcataa tgtaaataat 2280
aatttgtata ttgtgttgca gatgaaagta ttccaggaac agtgaatggg agaagacaca 2340
agaacatttg tttgtttgtc ttctgatgtt ttttcttaaa atagtaattt ctctactttt 2400
tcttttctac tgttgtctta actacagggtg attggaatgc caaacactct taagttttatt 2460
ttcttttttc gttttataaa ttcagtgtgc caaatgaaac ttttttccta agtaactgta 2520
ataggaaaaa gtttattttg agagttttctt cttcataaat ctacagacat taaacaattg 2580
ttgtgttctt tttacctttt atttttctat taccttgcta ccaaacagtt tagatagcaa 2640
tataatagca aaaaagcaaa tatggtaaaa tagagaagggt ttgaagggtt gagttactct 2700
gtcatataac atgtagatca gtcttcatgt gacctgcagt attttttttt ctaatgtatt 2760
tgtcagaaat ctgttgtaga ctgttaactt ctctctgatg gaattttatt tctgcaagaa 2820
ttattctgat atttaagaga gccaatttta actgctgtga aaatgtttcc agtgcaagag 2880
aagggaataa ctaggaacta agacatttct aatttattgc ttattacttt cttaatttta 2940
caggataatt ataagcaagt ggaactacca tcttttattc ttaataatta ttaatccctt 3000
caatgaaact ttaaaaaaac tgaattttta tacatggcat acatttttct agttccttct 3060
gcttgcttta ttaactcaaa agttctagtt ctagtctgtt gatctgcctt ttgttctccc 3120
aaaatgtaca gtaattccat ttgtttgtat aaatatgcct ggattttcat tataaaaatg 3180
tcattgtagg gagtagagac tcatatcatg gccttttaaa tattgtaata aaggcaaata 3240
gatatttgcc cttagtttac tgg 3263

```

<210> 3

<211> 616

<212> PRT

<213> Homo sapiens

<400> 3

Met Asn Ser Pro Gly Gly Arg Gly Lys Lys Lys Gly Ser Gly Gly Ala
1 5 10 15

Ser Asn Pro Val Pro Pro Arg Pro Pro Pro Pro Cys Leu Ala Pro Ala
20 25 30

Pro Pro Ala Ala Gly Pro Ala Pro Pro Pro Glu Ser Pro His Lys Arg
35 40 45

Asn	Leu	Tyr	Tyr	Phe	Ser	Tyr	Pro	Leu	Phe	Val	Gly	Phe	Ala	Leu	Leu
50						55					60				
Arg	Leu	Val	Ala	Phe	His	Leu	Gly	Leu	Leu	Phe	Val	Trp	Leu	Cys	Gln
65					70					75					80
Arg	Phe	Ser	Arg	Ala	Leu	Met	Ala	Ala	Lys	Arg	Ser	Ser	Gly	Ala	Ala
				85					90					95	
Pro	Ala	Pro	Ala	Ser	Ala	Ser	Ala	Pro	Ala	Pro	Val	Pro	Gly	Gly	Gln
			100					105					110		
Ala	Glu	Arg	Val	Arg	Val	Phe	His	Lys	Gln	Ala	Phe	Glu	Tyr	Ile	Ser
		115					120					125			
Ile	Ala	Leu	Arg	Ile	Asp	Glu	Asp	Glu	Lys	Ala	Gly	Gln	Lys	Glu	Gln
	130					135					140				
Ala	Val	Glu	Trp	Tyr	Lys	Lys	Gly	Ile	Glu	Glu	Leu	Glu	Lys	Gly	Ile
145					150					155					160
Ala	Val	Ile	Val	Thr	Gly	Gln	Gly	Glu	Gln	Cys	Glu	Arg	Ala	Arg	Arg
				165					170					175	
Leu	Gln	Ala	Lys	Met	Met	Thr	Asn	Leu	Val	Met	Ala	Lys	Asp	Arg	Leu
			180					185					190		
Gln	Leu	Leu	Glu	Lys	Met	Gln	Pro	Val	Leu	Pro	Phe	Ser	Lys	Ser	Gln
		195					200					205			
Thr	Asp	Val	Tyr	Asn	Asp	Ser	Thr	Asn	Leu	Ala	Cys	Arg	Asn	Gly	His
	210					215					220				
Leu	Gln	Ser	Glu	Ser	Gly	Ala	Val	Pro	Lys	Arg	Lys	Asp	Pro	Leu	Thr
225					230					235					240
His	Thr	Ser	Asn	Ser	Leu	Pro	Arg	Ser	Lys	Thr	Val	Met	Lys	Thr	Gly
				245					250					255	
Ser	Ala	Gly	Leu	Ser	Gly	His	His	Arg	Ala	Pro	Ser	Tyr	Ser	Gly	Leu
			260					265					270		
Ser	Met	Val	Ser	Gly	Val	Lys	Gln	Gly	Ser	Gly	Pro	Ala	Pro	Thr	Thr
		275					280					285			
His	Lys	Gly	Thr	Pro	Lys	Thr	Asn	Arg	Thr	Asn	Lys	Pro	Ser	Thr	Pro
	290					295					300				
Thr	Thr	Ala	Thr	Arg	Lys	Lys	Lys	Asp	Leu	Lys	Asn	Phe	Arg	Asn	Val
305					310					315					320
Asp	Ser	Asn	Leu	Ala	Asn	Leu	Ile	Met	Asn	Glu	Ile	Val	Asp	Asn	Gly
				325					330					335	
Thr	Ala	Val	Lys	Phe	Asp	Asp	Ile	Ala	Gly	Gln	Asp	Leu	Ala	Lys	Gln
			340					345					350		

Ala Leu Gln Glu Ile Val Ile Leu Pro Ser Leu Arg Pro Glu Leu Phe
355 360 365

Thr Gly Leu Arg Ala Pro Ala Arg Gly Leu Leu Leu Phe Gly Pro Pro
370 375 380

Gly Asn Gly Lys Thr Met Leu Ala Lys Ala Val Ala Ala Glu Ser Asn
385 390 395 400

Ala Thr Phe Phe Asn Ile Ser Ala Ala Ser Leu Thr Ser Lys Tyr Val
405 410 415

Gly Glu Gly Glu Lys Leu Val Arg Ala Leu Phe Ala Val Ala Arg Glu
420 425 430

Leu Gln Pro Ser Ile Ile Phe Ile Asp Glu Val Asp Ser Leu Leu Cys
435 440 445

Glu Arg Arg Glu Gly Glu His Asp Ala Ser Arg Arg Leu Lys Thr Glu
450 455 460

Phe Leu Ile Glu Phe Asp Gly Val Gln Ser Ala Gly Asp Asp Arg Val
465 470 475 480

Leu Val Met Gly Ala Thr Asn Arg Pro Gln Glu Leu Asp Glu Ala Val
485 490 495

Leu Arg Arg Phe Ile Lys Arg Val Tyr Val Ser Leu Pro Asn Glu Glu
500 505 510

Thr Arg Leu Leu Leu Lys Asn Leu Leu Cys Lys Gln Gly Ser Pro
515 520 525

Leu Thr Gln Lys Glu Leu Ala Gln Leu Ala Arg Met Thr Asp Gly Tyr
530 535 540

Ser Gly Ser Asp Leu Thr Ala Leu Ala Lys Asp Ala Ala Leu Gly Pro
545 550 555 560

Ile Arg Glu Leu Lys Pro Glu Gln Val Lys Asn Met Ser Ala Ser Glu
565 570 575

Met Arg Asn Ile Arg Leu Ser Asp Phe Thr Glu Ser Leu Lys Lys Ile
580 585 590

Lys Arg Ser Val Ser Pro Gln Thr Leu Glu Ala Tyr Ile Arg Trp Asn
595 600 605

Lys Asp Phe Gly Asp Thr Thr Val
610 615

<210> 4

<211> 23

<212> DNA

<213> Artificial sequence

<220>
 <223> Primer

 <400> 4
 cggagctcct cttggctgcc atg 23

 <210> 5
 <211> 26
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 5
 agaagcgctg gcagagccac acgaag 26

 <210> 6
 <211> 27
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 6
 aaggcgacca aacgcagcag cgcgaag 27

 <210> 7
 <211> 26
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 7
 aggagcaagc tgtggaatgg tataag 26

 <210> 8
 <211> 27
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 8
 tggttatggc caaggaccgc ttacaac 27

 <210> 9
 <211> 26
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

<400> 9
 caaacggacg tctataatga cagtac 26

 <210> 10
 <211> 25
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 10
 ttaggaatgt ggacagcaac cttgc 25

 <210> 11
 <211> 25
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 11
 cttctctgag gcctgagttg ttcac 25

 <210> 12
 <211> 27
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 12
 tgctagaatg actgatggat actcagg 27

 <210> 13
 <211> 24
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 13
 agatgcagca ctgggtccta tccg 24

 <210> 14
 <211> 26
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 14

26

<220>
<223> Primer

19

<220>
<223> Primer

20

<220>
<223> Primer

19

<220>
<223> Primer

20

<220>
<223> Primer

20

```
<210> 20
<211> 20
<212> DNA
<213> Artificial sequence
```

<220>
<223> Primer

```
<400> 20
ctgtgaacaa ctcaggcctc                20
```

```
<210> 21
<211> 20
<212> DNA
<213> Artificial sequence
```

<220>
<223> Primer

<400> 21
atgagaaagc aggacagaag 20

```
<210> 22
<211> 18
<212> DNA
<213> Artificial sequence
```

<220>
<223> Primer

<400> 22	
tgccaagtct tgaccagc	18

```
<210> 23
<211> 20
<212> DNA
<213> Artificial sequence
```

<220>
<223> Primer

<400> 23
ctacaactgc tactcgtaag 20

```
<210> 24
<211> 19
<212> DNA
<213> Artificial sequence
```

<220>
<223> Primer

<400> 24
cagtgcctgca tctttttgcc 19

```
<210> 25
<211> 20
```

<212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 25
 taggaatgtg gacagcaacc 20

<210> 26
 <211> 20
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 26
 aaagctgtta ggtcacttcc 20

<210> 27
 <211> 20
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 27
 tggagatgac agagtacttg 20

<210> 28
 <211> 19
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 28
 ctggaatact ttcactctgc 19

<210> 29
 <211> 19
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 29
 atgaggctgt tctcaggcg 19

<210> 30
 <211> 20
 <212> DNA
 <213> Artificial sequence

136707-2359359

<220>
 <223> Primer

 <400> 30
 gtgagccgaa ctgcacattg 20

 <210> 31
 <211> 21
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 31
 caaagtcgac agctacagtg c 21

 <210> 32
 <211> 20
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 32
 ggaactgtag ttgagtggga 20

 <210> 33
 <211> 19
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 33
 agatgaggct ccgacctac 19

 <210> 34
 <211> 20
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 34
 aatgccacac ttgtaatctc 20

 <210> 35
 <211> 22
 <212> DNA
 <213> Artificial sequence

 <220>

<223> Primer

<400> 35

tgtgaatata tcataatttg gg

22

<210> 36

<211> 19

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 36

tacagcagtt ctcatgatg

19

<210> 37

<211> 21

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 37

gaccaaattg gtgcatgcat g

21

<210> 38

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 38

acatttccaa tacatcccac

20

<210> 39

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 39

atttgtcatt tcacatgcac

20

<210> 40

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 40	
ttagaatgac tatacctgac	20
<210> 41	
<211> 18	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 41	
tcaggttaag taagactc	18
<210> 42	
<211> 19	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 42	
ttcctatcta cctagtgc	19
<210> 43	
<211> 20	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 43	
ttttatagca agttgccctg	20
<210> 44	
<211> 19	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 44	
cctatgaaga tcttggtac	19
<210> 45	
<211> 20	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 45	
tgatcatgatt ctaacaaggg	20

<210> 51

<211> 20
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 51
 cattgattgc catgtattgg 20

<210> 52
 <211> 20
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 52
 agaaggccag aaatactcag 20

<210> 53
 <211> 22
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 53
 gtactttaat cggtaaatat gg 22

<210> 54
 <211> 20
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 54
 ctcaagtctt aggaatgcag 20

<210> 55
 <211> 20
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer

<400> 55
 gcacttaacc aggctgtatg 20

<210> 56
 <211> 19
 <212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 56

ctcagatgac tcacatagc

19

<210> 57

<211> 22

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 57

ctttactaga ctaattctcc tg

22

<210> 58

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 58

cagattcaag aagacagatc

20

<210> 59

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 59

gcaataattc accacacttg

20

<210> 60

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> Primer

<400> 60

ggtagttctt gtttctgctc

20

<210> 61

<211> 20

<212> DNA

<213> Artificial sequence

106072953350

<220>
 <223> Primer

 <400> 61
 caagtgtggt gaattattgc 20

 <210> 62
 <211> 19
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 62
 gagctgaaaa gtattcagc 19

 <210> 63
 <211> 20
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 63
 tgcaaaggac atagccagtg 20

 <210> 64
 <211> 20
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 64
 agcctctgga gatagtatgc 20

 <210> 65
 <211> 20
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

 <400> 65
 ctagaacagg ggtcacagtc 20

 <210> 66
 <211> 18
 <212> DNA
 <213> Artificial sequence

 <220>
 <223> Primer

105407-2050250

<400> 66	
ttggacttct taaacttc	18
<210> 67	
<211> 21	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 67	
gcagtatgca agaaattgaa c	21
<210> 68	
<211> 19	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 68	
ggcctgtaat tttcttctg	19
<210> 69	
<211> 21	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 69	
gtactgaata gatacatgta g	21
<210> 70	
<211> 19	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 70	
gtgtagcaga tcaacatag	19
<210> 71	
<211> 20	
<212> DNA	
<213> Artificial sequence	
<220>	
<223> Primer	
<400> 71	

catcttcaag tttggtgcac

20

<210> 72
<211> 1689
<212> DNA
<213> mouse

<220>
<223> Incomplete Spg4 DNA.

<400> 72

aggccgagag cgtccgcgctc ttccacaagc aggccttcga gtacatctcc attgccctgc 60
gcatcgacga ggaagagaaa gcaggacaga aggaacaagc tgtggaatgg tataagaaag 120
gtatcgaaga actggaaaaa ggaatcgctg ttatagttac gggccaaggt gaacagtatg 180
aaagagctag acgtcttcaa gcaaaaatga tgactaattt agttatggcc aaggaccgtt 240
tacaacttct agagaagctg caaccagttt tgcaattttc caagtcacag acggacgtct 300
ataacgagag tactaacctg acatgccgca atggacatct ccagtcagaa agtggagcag 360
ttccgaagag gaaagacccc ttaacacatg ctagtaattc attgcctcga tcaaaaactg 420
tcctgaaaag tggtccgca gggctctccg gtcaccacag ggcgccagt tgcagtgggt 480
tgtccatggt ttctggagca agaccgggac ctggtcctgc agctaccaca cataagggta 540
ctccaaaacc aaatagaacc acaaaacctt ctactccac aactgcagtt cggaaaaaga 600
aagacttgaa aaatttttagg aatgtggaca gcaatcttgc taaccttata atgaatgaaa 660
ttgttgacaa tgggacagct gttaagtttg atgacatagc cgggcaggag ctggcaaagc 720
aagcgctgca ggagattgtc atccttcctt ctctgcggcc tgagttgttc acagggctca 780
gagctcctgc tagaggcttg ttactcttcg gtccgccagg aaacggaaaa acaatgctgg 840
ctaaagcagt agctgcagag tctaattgca cctttttcaa cataagtgtc gccagtttaa 900
cttcaaaaata tgtgggagaa ggagagaaat tggtagagc tctctttgct gtggctcgag 960
aacttcaacc atctataatt tttatagatg aagttgacag tcttttgtgt gagagacggg 1020
aaggggagca cgacgctagc agacggctaa agacggaatt tttaatagaa tttgacgggg 1080
tgcaatctgc tggagatgac agagtacttg taatgggtgc aactaacagg cccaagagc 1140
ttgatgaagc tgttctcagg cgtttcatta aacgggtata tgtgtcctta ccaaatgagg 1200
agacaagact ccttctgctt aaaaacctgt tgtgtaaaca aggaagtcca ctgacccaaa 1260
aagaactcgc acagcttgct agaattgacc atggatactc tgggaagtgt ctgaccgctt 1320
tggccaagga tgcagccctg ggtcctatcc gagaactgaa gccagagcag gtgaagaata 1380
tgtctgccag tgagatgaga aatattcgat tatctgactt cacagaatcc ttaaaaaaga 1440
taaaacgcag tgtgagtcct cagaccttag aagcatacat acgctggaac aaggattttg 1500
gagacaccac tgttttaaagg aatggatgcc tctgtgagcc catagaacat cgcacttcac 1560
aggaacaag agctttggct acaggaaccc agacttcgtt tacaggacgt tttagagttt 1620
tcatttttgt gcaccaaact tgaagaggaa caagaagaca gacctaaata aaatatgcaa 1680
tatgaatgg 1689

<210> 73
<211> 504
<212> PRT
<213> mouse

<220>
<223> Incomplete murine spastin.

<400> 73

Ala	Glu	Ser	Val	Arg	Val	Phe	His	Lys	Gln	Ala	Phe	Glu	Tyr	Ile	Ser
1				5					10					15	
Ile	Ala	Leu	Arg	Ile	Asp	Glu	Glu	Glu	Lys	Ala	Gly	Gln	Lys	Glu	Gln
			20					25					30		

Ala	Val	Glu	Trp	Tyr	Lys	Lys	Gly	Ile	Glu	Glu	Leu	Glu	Lys	Gly	Ile
		35					40					45			
Ala	Val	Ile	Val	Thr	Gly	Gln	Gly	Glu	Gln	Tyr	Glu	Arg	Ala	Arg	Arg
	50					55					60				
Leu	Gln	Ala	Lys	Met	Met	Thr	Asn	Leu	Val	Met	Ala	Lys	Asp	Arg	Leu
65					70					75					80
Gln	Leu	Leu	Glu	Lys	Leu	Gln	Pro	Val	Leu	Gln	Phe	Ser	Lys	Ser	Gln
				85					90					95	
Thr	Asp	Val	Tyr	Asn	Glu	Ser	Thr	Asn	Leu	Thr	Cys	Arg	Asn	Gly	His
			100					105					110		
Leu	Gln	Ser	Glu	Ser	Gly	Ala	Val	Pro	Lys	Arg	Lys	Asp	Pro	Leu	Thr
		115					120					125			
His	Ala	Ser	Asn	Ser	Leu	Pro	Arg	Ser	Lys	Thr	Val	Leu	Lys	Ser	Gly
	130					135					140				
Ser	Ala	Gly	Leu	Ser	Gly	His	His	Arg	Ala	Pro	Ser	Cys	Ser	Gly	Leu
145					150					155					160
Ser	Met	Val	Ser	Gly	Ala	Arg	Pro	Gly	Pro	Gly	Pro	Ala	Ala	Thr	Thr
				165				170						175	
His	Lys	Gly	Thr	Pro	Lys	Pro	Asn	Arg	Thr	Asn	Lys	Pro	Ser	Thr	Pro
			180					185					190		
Thr	Thr	Ala	Val	Arg	Lys	Lys	Lys	Asp	Leu	Lys	Asn	Phe	Arg	Asn	Val
		195					200					205			
Asp	Ser	Asn	Leu	Ala	Asn	Leu	Ile	Met	Asn	Glu	Ile	Val	Asp	Asn	Gly
	210					215					220				
Thr	Ala	Val	Lys	Phe	Asp	Asp	Ile	Ala	Gly	Gln	Glu	Leu	Ala	Lys	Gln
225					230					235					240
Ala	Leu	Gln	Glu	Ile	Val	Ile	Leu	Pro	Ser	Leu	Arg	Pro	Glu	Leu	Phe
				245					250					255	
Thr	Gly	Leu	Arg	Ala	Pro	Ala	Arg	Gly	Leu	Leu	Leu	Phe	Gly	Pro	Pro
			260					265					270		
Gly	Asn	Gly	Lys	Thr	Met	Leu	Ala	Lys	Ala	Val	Ala	Ala	Glu	Ser	Asn
		275					280					285			
Ala	Thr	Phe	Phe	Asn	Ile	Ser	Ala	Ala	Ser	Leu	Thr	Ser	Lys	Tyr	Val
	290					295					300				
Gly	Glu	Gly	Glu	Lys	Leu	Val	Arg	Ala	Leu	Phe	Ala	Val	Ala	Arg	Glu
305					310					315					320
Leu	Gln	Pro	Ser	Ile	Ile	Phe	Ile	Asp	Glu	Val	Asp	Ser	Leu	Leu	Cys
				325					330					335	

Glu Arg Arg Glu Gly Glu His Asp Ala Ser Arg Arg Leu Lys Thr Glu
 340 345 350
 Phe Leu Ile Glu Phe Asp Gly Val Gln Ser Ala Gly Asp Asp Arg Val
 355 360 365
 Leu Val Met Gly Ala Thr Asn Arg Pro Gln Glu Leu Asp Glu Ala Val
 370 375 380
 Leu Arg Arg Phe Ile Lys Arg Val Tyr Val Ser Leu Pro Asn Glu Glu
 385 390 395 400
 Thr Arg Leu Leu Leu Leu Lys Asn Leu Leu Cys Lys Gln Gly Ser Pro
 405 410 415
 Leu Thr Gln Lys Glu Leu Ala Gln Leu Ala Arg Met Thr Asp Gly Tyr
 420 425 430
 Ser Gly Ser Asp Leu Thr Ala Leu Ala Lys Asp Ala Ala Leu Gly Pro
 435 440 445
 Ile Arg Glu Leu Lys Pro Glu Gln Val Lys Asn Met Ser Ala Ser Glu
 450 455 460
 Met Arg Asn Ile Arg Leu Ser Asp Phe Thr Glu Ser Leu Lys Lys Ile
 465 470 475 480
 Lys Arg Ser Val Ser Pro Gln Thr Leu Glu Ala Tyr Ile Arg Trp Asn
 485 490 495
 Lys Asp Phe Gly Asp Thr Thr Val
 500

<210> 74
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice acceptor site.

<400> 74
 attttttatt tttaaagcagg acag

24

<210> 75
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice acceptor site.

```
<210> 76
<211> 24
<212> DNA
<213> Homo sapiens
```

<220>
<223> SPG4 gene splice acceptor site.

<400> 76
cttctctgtt gcatagagaa gatg 24

```
<210> 77
<211> 24
<212> DNA
<213> Homo sapiens
```

<220>
<223> SPG4 gene splice acceptor site.

<400> 77
acttttttctt tgtcagaaag tggg 24

```
<210> 78
<211> 24
<212> DNA
<213> Homo sapiens
```

<220>
<223> SPG4 gene splice acceptor site.

<400> 78
 ttttgtatcc ttttaagggtta ctcc 24

```
<210> 79
<211> 24
<212> DNA
<213> Homo sapiens
```

```
<220>
<223> SPG4 gene splice acceptor site.
```

<400> 79
aggtcttgggt tcttagtgga acag 24

```
<210> 80
<211> 24
<212> DNA
<213> Homo sapiens
```

<220>

<223> SPG4 gene splice acceptor site.

<400> 80

agtatatatt ttttagttgt tcac

24

<210> 81

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<223> SPG4 gene splice acceptor site.

<400> 81

cttgtgattt ttaaaggcta aagc

24

<210> 82

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<223> SPG4 gene splice acceptor site.

<400> 82

taatgctttg ttttaggtgg gaga

24

<210> 83

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<223> SPG4 gene splice acceptor site.

<400> 83

cttgtatttc ctctagatga agtt

24

<210> 84

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<223> SPG4 gene splice acceptor site.

<400> 84

gattttttgc ttgtaggtac agtc

24

<210> 85

<211> 24

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

<212> DNA
<213> Homo sapiens

<220>
<223> SPG4 gene splice acceptor site.

<400> 85
ggattttttt ttttaggcgt ttca 24

<210> 86
<211> 24
<212> DNA
<213> Homo sapiens

<220>
<223> SPG4 gene splice acceptor site.

<400> 86
ttttaatatt tttcagacaa gact 24

<210> 87
<211> 24
<212> DNA
<213> Homo sapiens

<220>
<223> SPG4 gene splice acceptor site.

<400> 87
tccttcctt cctcagaatg actg 24

<210> 88
<211> 24
<212> DNA
<213> Homo sapiens

<220>
<223> SPG4 gene splice acceptor site.

<400> 88
cttttatgtt ttacagaact aaaa 24

<210> 89
<211> 24
<212> DNA
<213> Homo sapiens

<220>
<223> SPG4 gene splice acceptor site.

<400> 89
ctttttaaaa atctagatga gaaa 24

<210> 90
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

<400> 90
 tgagaaaggt aactaggggg ctgg 24

<210> 91
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

<400> 91
 aggacaaggt aagattgtat ttgt 24

<210> 92
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

<400> 92
 acttctaggt atcaattaat gtat 24

<210> 93
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

<400> 93
 ccagtcaggt gggtttaggt taac 24

<210> 94
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

```
<210> 99
<211> 24
<212> DNA
<213> Homo sapiens
```

<220>
 <223> SPG4 gene splice donor site.

<400> 99
 ttttataggt aagaacatat tttc 24

<210> 100
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

<400> 100
 ttgatggtgt aagtgttgat tatg 24

<210> 101
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

<400> 101
 gttctcaggt agggagattt atat 24

<210> 102
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

<400> 102
 atgaggaggt atgtatctgt gttt 24

<210> 103
 <211> 24
 <212> DNA
 <213> Homo sapiens

<220>
 <223> SPG4 gene splice donor site.

<400> 103
 cttgctaggt gagtaatttg gatt 24

<210> 104
 <211> 24
 <212> DNA
 <213> Homo sapiens

SPG4 gene splice donor site.

<220>

<223> SPG4 gene splice donor site.

<400> 104

tatccgaggt aggtatacaa gagg

24

<210> 105

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<223> SPG4 gene splice donor site.

<400> 105

ccagtgaggt atagtatttt acaa

24

<210> 106

<211> 2152

<212> DNA

<213> Mus musculus

<220>

<223> Complete Spg4 DNA

<400> 106

```

gtccccgggc tccggcgggc gcgcgagcgg ctccgtgggc ccccgccgcc gcagtggcag 60
tgccgcgcgc cgcgcgttgg tcgccgtcgg tctgcgggaa gcgggttatg gcggcggcgg 120
cagtgggagc tgtgaatgag ttctccggcc ggacgacgga agaagaaagg ctcgggcggc 180
gcgagcccg cgcgcgccag gcctccggcc cccgcgcgg tccccgcccc tgccgcgcgc 240
ccggccccctg cggccggctc gccgcctaag cggaacctgt cttctttctc gtccccgctg 300
gtcgtcggct tcgccctgct gcgcctgctg gectgccacc tggggctcct ctccgcgtgg 360
ctctgccagc gtttctcccg cgcctcatg gccgccaaga ggagctccgg gaccgcgcgc 420
gcgcccgcct cgcctcgcgc cccagcgcgc ggaccgggtg gcgaggccga gacggtccgc 480
gtcttccaca agcaggcctt cgagtacatc tccattgccc tgcgcacatga cgagggaagag 540
aaagcaggac agaaggaaca agctgtggaa tggataaaga aaggtatcga agaactggaa 600
aaaggaatcg ctgttatagt tacgggcca ggtgaacagt atgaaagagc tagacgtctt 660
caagccaaaa tgatgactaa tttagttatg gccaaaggacc gtttacaact tctagagaag 720
ctgcaaccag ttttgcaatt ttccaagtca cagacggacg tctataacga gagtactaac 780

```

```

ctgacatgcc gcaatggaca tctccagtca gaaagtggag cagttccgaa gaggaaagac 840
cccttaacac atgctagtaa ttcattgcct cgatcaaaaa ctgtcctgaa aagtggctcc 900
gcagggtctt ccggtcacca cagggcgcct agttgcagtg gtttgtccat ggtttctgga 960
gcaagaccgg gacctggtcc tgcagctacc acacataagg gtactccaaa accaaataga 1020
accaacaaac cttctactcc cacaactgca gttcggaaaa agaaagactt gaaaaatatt 1080
aggaatgtgg acagcaatct tgctaaccctt ataataatg aaattggtga caatgggaca 1140
gctgttaagt ttgatgacat agccgggcag gagctggcaa agcaagcgtc gcaggagatt 1200
gtcatccttc cttctctgcg gcctgagttg ttcacagggc tcagagctcc tgctagaggc 1260
ttgttactct tcgggtccgc aggaacgga aaaacaatgc tggctaaagc agtagctgca 1320
gagtctaatt cgaccttttt caacataagt gctgccagtt taacttcaaa atatgtggga 1380
gaaggagaga aattgggtgag agctctcttt gctgtggctc gagaacttca accatctata 1440
atttttatag atgaagttga cagtcttttg tgtgagagac gggaaagggga gcacgacgct 1500
agcagacggc taaagacgga atttttaata gaatttgacg ggggtgcaatc tgctggagat 1560
gacagagtac ttgtaatggg tgcaactaac aggcccaag agcttgatga agctgttctc 1620

```

```
<210> 107
<211> 614
<212> PRT
<213> Mus musculus
```

<400> 107															
Met 1	Ser	Ser	Pro	Ala 5	Gly	Arg	Arg	Lys	Lys 10	Lys	Gly	Ser	Gly	Gly 15	Ala
Ser	Pro	Ala	Pro 20	Ala	Arg	Pro	Pro	Pro 25	Pro	Ala	Ala	Val	Pro 30	Ala	Pro
Ala	Ala	Gly 35	Pro	Ala	Pro	Ala	Ala 40	Gly	Ser	Pro	Pro	Lys 45	Arg	Asn	Leu
Ser	Ser 50	Phe	Ser	Ser	Pro	Leu 55	Val	Val	Gly	Phe	Ala 60	Leu	Leu	Arg	Leu
Leu 65	Ala	Cys	His	Leu	Gly 70	Leu	Leu	Phe	Ala 75	Trp	Leu	Cys	Gln	Arg	Phe 80
Ser	Arg	Ala	Leu	Met 85	Ala	Ala	Lys	Arg	Ser 90	Ser	Gly	Thr	Ala	Pro 95	Ala
Pro	Ala	Ser 100	Pro	Ser	Pro	Pro	Ala 105	Pro	Gly	Pro	Gly	Gly 110	Glu	Ala	Glu
Ser	Val	Arg 115	Val	Phe	His	Lys	Gln 120	Ala	Phe	Glu	Tyr	Ile 125	Ser	Ile	Ala
Leu 130	Arg	Ile	Asp	Glu	Glu 135	Glu	Lys	Ala	Gly	Gln	Lys 140	Glu	Gln	Ala	Val
Glu 145	Trp	Tyr	Lys	Lys 150	Gly	Ile	Glu	Glu	Leu 155	Glu	Lys	Gly	Ile	Ala	Val 160
Ile	Val	Thr	Gly 165	Gln	Gly	Glu	Gln	Tyr	Glu 170	Arg	Ala	Arg	Arg	Leu 175	Gln
Ala	Lys	Met 180	Met	Thr	Asn	Leu	Val 185	Met	Ala	Lys	Asp	Arg	Leu 190	Gln	Leu
Leu	Glu	Lys	Leu	Gln	Pro	Val	Leu	Gln	Phe	Ser	Lys	Ser	Gln	Thr	Asp

195					200					205					
Val	Tyr	Asn	Glu	Ser	Thr	Asn	Leu	Thr	Cys	Arg	Asn	Gly	His	Leu	Gln
	210					215					220				
Ser	Glu	Ser	Gly	Ala	Val	Pro	Lys	Arg	Lys	Asp	Pro	Leu	Thr	His	Ala
225					230					235					240
Ser	Asn	Ser	Leu	Pro	Arg	Ser	Lys	Thr	Val	Leu	Lys	Ser	Gly	Ser	Ala
				245					250					255	
Gly	Leu	Ser	Gly	His	His	Arg	Ala	Pro	Ser	Cys	Ser	Gly	Leu	Ser	Met
			260					265					270		
Val	Ser	Gly	Ala	Arg	Pro	Gly	Pro	Gly	Pro	Ala	Ala	Thr	Thr	His	Lys
		275					280					285			
Gly	Thr	Pro	Lys	Pro	Asn	Arg	Thr	Asn	Lys	Pro	Ser	Thr	Pro	Thr	Thr
	290					295					300				
Ala	Val	Arg	Lys	Lys	Lys	Asp	Leu	Lys	Asn	Phe	Arg	Asn	Val	Asp	Ser
305					310					315					320
Asn	Leu	Ala	Asn	Leu	Ile	Met	Asn	Glu	Ile	Val	Asp	Asn	Gly	Thr	Ala
				325					330					335	
Val	Lys	Phe	Asp	Asp	Ile	Ala	Gly	Gln	Glu	Leu	Ala	Lys	Gln	Ala	Leu
			340					345					350		
Gln	Glu	Ile	Val	Ile	Leu	Pro	Ser	Leu	Arg	Pro	Glu	Leu	Phe	Thr	Gly
	355						360					365			
Leu	Arg	Ala	Pro	Ala	Arg	Gly	Leu	Leu	Leu	Phe	Gly	Pro	Pro	Gly	Asn
	370					375					380				
Gly	Lys	Thr	Met	Leu	Ala	Lys	Ala	Val	Ala	Ala	Glu	Ser	Asn	Ala	Thr
385					390					395					400
Phe	Phe	Asn	Ile	Ser	Ala	Ala	Ser	Leu	Thr	Ser	Lys	Tyr	Val	Gly	Glu
			405						410					415	
Gly	Glu	Lys	Leu	Val	Arg	Ala	Leu	Phe	Ala	Val	Ala	Arg	Glu	Leu	Gln
			420					425					430		
Pro	Ser	Ile	Ile	Phe	Ile	Asp	Glu	Val	Asp	Ser	Leu	Leu	Cys	Glu	Arg
		435					440					445			
Arg	Glu	Gly	Glu	His	Asp	Ala	Ser	Arg	Arg	Leu	Lys	Thr	Glu	Phe	Leu
	450					455					460				
Ile	Glu	Phe	Asp	Gly	Val	Gln	Ser	Ala	Gly	Asp	Asp	Arg	Val	Leu	Val
465					470					475					480
Met	Gly	Ala	Thr	Asn	Arg	Pro	Gln	Glu	Leu	Asp	Glu	Ala	Val	Leu	Arg
				485					490					495	
Arg	Phe	Ile	Lys	Arg	Val	Tyr	Val	Ser	Leu	Pro	Asn	Glu	Glu	Thr	Arg

500					505					510					
Leu	Leu	Leu	Leu	Lys	Asn	Leu	Leu	Cys	Lys	Gln	Gly	Ser	Pro	Leu	Thr
		515					520					525			
Gln	Lys	Glu	Leu	Ala	Gln	Leu	Ala	Arg	Met	Thr	Asp	Gly	Tyr	Ser	Gly
	530					535					540				
Ser	Asp	Leu	Thr	Ala	Leu	Ala	Lys	Asp	Ala	Ala	Leu	Gly	Pro	Ile	Arg
545						550					555				560
Glu	Leu	Lys	Pro	Glu	Gln	Val	Lys	Asn	Met	Ser	Ala	Ser	Glu	Met	Arg
				565					570					575	
Asn	Ile	Arg	Leu	Ser	Asp	Phe	Thr	Glu	Ser	Leu	Lys	Lys	Ile	Lys	Arg
			580					585					590		
Ser	Val	Ser	Pro	Gln	Thr	Leu	Glu	Ala	Tyr	Ile	Arg	Trp	Asn	Lys	Asp
		595					600					605			
Phe	Gly	Asp	Thr	Thr	Val										
	610														